Logical radio maps for user localization in a real outdoor radio environment

Pere Garau Burguera

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 23.11.2020

Supervisor

Prof. Olav Tirkkonen

Advisor

Dr. Hanan Al-Tous

Aalto University
School of Electrical
Engineering

Aalto University
School of Electrical
Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Pere Garau Burguera

Title Logical radio maps for user localization in a real outdoor radio environment

Degree programme Computer, Communication and Information Sciences

Major Communications Engineering **Code of major** ELEC3029

Supervisor Prof. Olav Tirkkonen

Advisor Dr. Hanan Al-Tous

Date 23.11.2020 **Number of pages** 68 **Language** English

Abstract

Channel charting (CC) is a novel framework in which a multi-antenna base station learns a channel chart, which is a low-dimensional representation of the true locations of transmitting users. This is done with the use of high-dimensional Channel State Information (CSI) acquired at the BS, and it operates in an unsupervised manner. The charting function, which maps high-dimensional features extracted from the CSI, to a channel chart, does not have any information about the locations of users. The channel chart has the property of local geometry preservation, i.e., points which are close in the physical space will also be close in the channel chart. In this thesis, the objective is to investigate the applications of CC in locating users within the radio environment. Using CSI collected from a *real* outdoor environment, it is attempted to estimate the unknown true locations of transmitters, from the information conveyed by channel charts learned from relevant features extracted from the CSI. Information about the antenna array at the BS is not used when processing the CSI. Different covariance matrix based features and their dissimilarity measures are investigated for the CC problem, and a wide range of state-of-the-art Dimensionality Reduction methods are applied. The use of *real* measurements, as opposed to simulated data, introduces new challenges in the prospect of CC and its user localization applications. Several of these challenges are considered in this thesis, including incomplete information about the antenna structure, small number of time samples or poor received signal. The results obtained are remarkable in that, despite the challenges mentioned, it is possible to obtain channel charts of high accuracy, and the performance of localization based on CC approaches is similar to that obtained based on high-dimensional features, with the benefit of the low-dimensionality of the channel charts.

Keywords MIMO, Channel State Information, Radio features, User localization, Channel Charting, Dimensionality Reduction, Machine learning

Preface

I want to thank Prof. Olav Tirkkonen for his wise guidance and support all along the development of this work, and for giving me the opportunity of working as a research assistant in his group, and continuing as a PhD student. I would also like to thank my advisor Dr. Hanan Al-Tous for her extensive commitment, suggestions and feedback, and also for introducing this topic to me. Her valuable insights have taught me useful skills and practices related to research.

I would like to give thanks to the nice friends I have met here in Finland, without whom nothing would have been the same. My gratitude extends to my old time friends back in Spain.

Last, but not least, I would like to express my gratitude to my family, for their sincere support along the way.

Otaniemi, 23.11.2020

Pere Garau Burguera

Contents

Symbols and abbreviations

Symbols

N, n	number of transmitters, transmitter index
A, a	number of antennas, antenna index
F, f	number of subcarriers, subcarrier index
S, s	number of samples, sample index
$\mathbf{x}$	(column) vector
$\mathbf{A}$	matrix
$[\mathbf{A}]_{ij}$	element of matrix $\mathbf{A}$ on the i-th row and j-th column
$\mathcal{X}$	set
$h_{n,a,f,s}$	channel response associated to transmitter n, antenna a, subcarrier f, and sample s
$\mathbf{R}_n$	covariance matrix related to transmitter n
$\bar{P}_n$	average power received from an UE at location n
$\overline{\mathrm{SNR}}_n$	average SNR related to location n
$\mathbf{I}_d$	identity matrix of size d
$\mathbf{1}$	vector of ones
$\mathbb{R}_+$	set of real numbers, including 0
$\mathbb{R}_+^*$	set of real numbers, excluding 0
K	number of neighbors
s'	dimension of the spatial geometry
s	dimension of the CSI
p	dimension of the features
p'	dimension of the channel chart
r	correlation coefficient
$\lambda, \mathbf{\Lambda}$	eigenvalue, matrix of eigenvalues
$\mathbf{v}, \mathbf{V}$	eigenvector, diagonal matrix of eigenvectors
$G = (\mathcal{V}, \mathcal{E})$	graph G, defined with a set of vertices, $\mathcal{V}$, and a set of edges connecting them, $\mathcal{E}$
$\mathbf{D}$	degree matrix of a graph
$\mathbf{W}$	weight matrix of a graph
$\mathbf{L}$	unnormalized Laplacian matrix of a graph
$\mathcal{L}$	normalized Laplacian matrix of a graph
$N(\mu, \sigma^2)$	real normal distribution with mean μ and variance σ^2
$CN(\mu, \sigma^2)$	complex normal distribution with mean μ and variance σ^2

Operators

$(\cdot)^H$	Hermitian conjugate of a vector/matrix
$(\cdot)^{-1}$	inverse of a square matrix
$(\cdot)^T$	transpose of a vector/matrix

$\|\cdot\|_F$	Frobenius norm of a matrix
$\|\cdot\|$	Euclidean norm of a vector
$\mathrm{Tr}(\cdot)$	trace of a matrix
$\mathrm{vec}(\cdot)$	elements of a matrix stacked in a column
$\mathbb{E}_x[\cdot]$	expectation, across variable x
$\Re(\cdot)$	real part
$\Im(\cdot)$	imaginary part
$\|\cdot\|$	magnitude of a complex number, cardinality of a set
$\angle(\cdot)$	argument of a complex number
$d_{\mathrm{diss}}(\cdot,\cdot)$	dissimilarity defined between two transmitter locations
$\mathrm{diag}(a_1,\ldots,a_\ell)$	matrix formed with elements $a_1,\ldots,a_\ell$ in its diagonal, with the rest as 0
$\mathrm{diag}(\cdot)$	vector formed from the diagonal elements of a matrix

Abbreviations

3GPP	Third Generation Partnership Project
5G	Fifth-Generation
AoA	Angle of Arrival
BS	Base Station
CC	Channel Charting
CDF	Cumulative Density Function
CMD	Correlation Matrix Distance
CSI	Channel State Information
CT	Continuity
DR	Dimensionality reduction
ELM	Extreme Learning Machine
GPS	Global Positioning System
KNN	K-Nearest Neighbors
KS	Kruskal's Stress
LE	Laplacian Eigenmaps
LLE	Locally Linear Embedding
LTSA	Local Tangent Space Alignment
LoS	Line of Sight
MDS	Multidimensional Scaling
MIMO	Multiple Input Multiple Output
mmWave	Millimeter Wave
MUSIC	Multiple Signal Classification
NLoS	Non-Line of Sight
OFDM	Orthogonal Frequency Division Multiplexing
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RF	Radio-Frequency
RRM	Radio Resource Management
RSS	Received Signal Strength

SNE	Stochastic Neighbor Embedding
SNR	Signal-to-Noise-Ratio
TDoA	Time Difference of Arrival
ToA	Time of Arrival
t-SNE	t-Distributed Stochastic Neighbor Embedding
TW	Trustworthiness
UE	User Equipment

1 Introduction

The wireless communication prospect is rapidly moving towards a scenario consisting of a massive increase in traffic volumes and number of connected devices, whose requirements—such as reliability and latency—are, in turn, becoming stricter. This situation poses complex challenges to the design of wireless communication systems. In order to cope with these challenges, several technologies have been object of research. In the level of radio access, some of these new technologies are massive Multiple-Input-Multiple-Output (mMIMO), the use of millimeter-Wave (mmWave) frequency bands and dense cell scenarios. These technologies are being deployed in wireless networks, especially in the scheme of Fifth-Generation (5G) networks.

In spite of the flagrant performance gain obtained from the deployment of these technologies, potential solutions, such as multiband operation [1] or mobile relaying [2], entail the acquisition of large amounts of multi-point Channel State Information (CSI) at multi-antenna Base Stations (BSs). One the one hand, networks can benefit from the vast amounts of CSI collected by BSs. On the other hand, storing and processing large amounts of CSI poses severe challenges in the design of wireless networks, which have to cope with large amounts of data.

For the purpose of effectively and efficiently managing the significant amounts of high-dimensional CSI, wireless networks must gain knowledge of the spatial distributions and trajectories of transmitting User Equipments (UEs), as well as their neighborhood relationships or handover procedures between neighboring cells. In other words, networks must learn the radio geometry in which the UEs reside. It is worth mentioning that a system where transmitters send their global coordinates to the BS would be highly inefficient, introducing significant resource consuming functions. Furthermore, this approach might intrude on the privacy of users, especially should the localization data be used and saved for other applications.

Channel Charting (CC), initially proposed in [3], is an unsupervised machine learning framework which aims at learning the high-dimensional radio geometry by learning a low-dimensional embedding, the so-called channel chart. In this channel chart, the local geometry of the location of UEs in space is preserved, relying solely on the CSI associated to each of the UEs, without the need of true location information obtained from positioning systems. The unsupervised nature of CC allows for a wide range of applications, including user localization, network planning, cell search, multi-point connectivity and handover, beam finding, user grouping, etc.

This thesis focuses on the utilization of CC for user localization. The fact that channel charts can be used for this purpose comes from the fact that they preserve the original local geometry. Localization is an essential function for wireless networks, and it is becoming more and more valuable for an endless number of applications, including autonomous driving, surveillance, disaster relief and pervasive sensor networks, among others.

In particular, the goal of this thesis is to explore and evaluate various localization methods from channel charts constructed from CSI collected at a real outdoor environment, while identifying the challenges of this task. Additionally, it aims to evaluate the charts obtained through various feature extraction and Dimensionality

Reduction (DR) techniques.

It is important to emphasize the nature of the dataset which is used in this thesis. The proposal of CC was based on synthetic data using a ray tracing model and a Third Generation Partnership Project (3GPP) based simulator [3]. On the other hand, the dataset used in this thesis contains real measurements from a real scenario, in which there is no control over the properties of the radio environment and the number of measurements taken. Moreover, the antennas and Radio-Frequency (RF) components of the transceivers are non-ideal, whereas a synthetic model usually simplifies the complexity by considering ideal components. Thus, this work aims to focus on the challenges related to the real nature of the dataset.

2 Background

This chapter includes a brief description of the relevant prior art related to user localization and charting methods in wireless environments. Additionally, a motivation for the use of channel charting in multi-user and multi-antenna scenarios is presented.

CC was originally proposed in [3] as a framework in which a MIMO base station is able to extract information from the high amounts of CSI collected, and learn, by means of a radio chart, the radio geometry in which the users are located. This chart is a representation of the spatial geometry, i.e. their true locations. As a consequence, charts can be leveraged, among other applications, for user localization; that is, estimating potentially unknown locations of users. The channel chart is constructed in a completely unsupervised manner, i.e., solely on features related to multi-antenna CSI, and therefore there is no need for the true locations to be a known by the BS.

The unsupervised nature of CC comprises its novelty, compared to other existing localization and user fingerprinting methods which have some degree of supervision. The motivation for the use of unsupervised frameworks is clear: the amount of CSI acquired at BSs depends on the number of UEs and its mobility, the bandwidth of operation, the number of receive and transmit antennas and the rate in which the wireless channel is estimated, among other factors. The amount of CSI needed for the correct functioning of a wireless network is thus growing, as user requirements become stricter. In order to efficiently and dynamically cope with high amounts of CSI, the construction of channel charts is unsupervised, which in addition automates this activity at the BS, by avoiding possible human intervention needed for supervised training.

Various research directions have been followed for the purpose of wireless user localization and fingerprinting. There has been a notorious effort spent on these problems. One approach is localization based on triangulation methods which rely on low-level parameters such as Angle-of-Arrival (AoA), Time-of-Arrival (ToA), Time-Difference-of-Arrival (TDoA) and Received Signal Strength (RSS) [4–6]. Supervised methods based on deep neural networks have been vastly studied [7–12]. Other approaches use manifold learning methods applied to user localization [13, 14]. Some localization approaches focus on semi-supervised metric learning from CSI, in which a metric with desired properties is defined, between users [15, 16].

CC is constructed following the steps depicted in Figure 1. CC is based on transforming multi-dimensional CSI, collected from multiple users and antennas, into a low-dimensional chart—typically 2-dimensional—in which its geometry has desired properties. Particularly, the interest lies in obtaining a chart in which the local geometry is preserved. That is, short distances should be preserved, therefore those users who are nearby in the spatial geometry, keep their neighborhood relationships in the channel chart.

There are four geometries which are referenced throughout this thesis. Wireless devices are said to reside in the spatial geometry, which encompasses their true locations and neighborhood relationships in the area of study. The CSI acquired

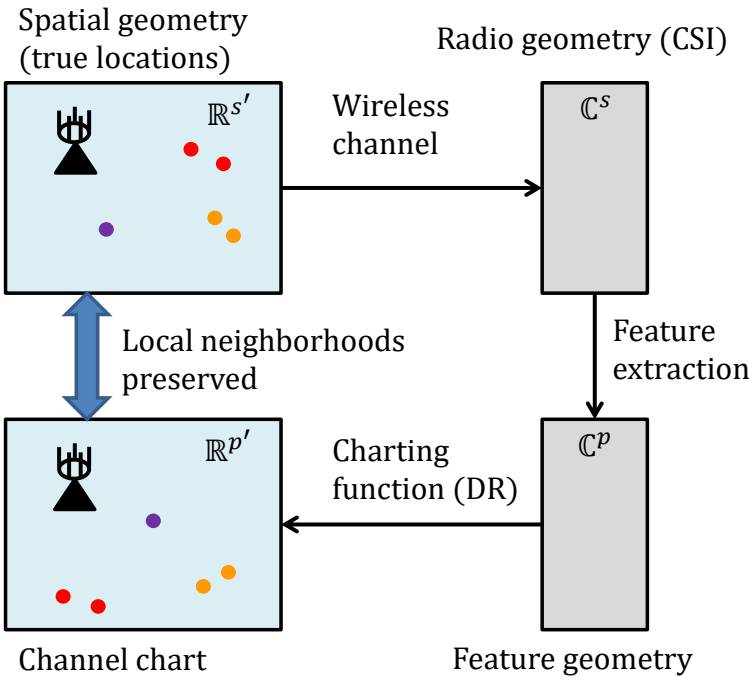


Figure 1: Diagram depicting the steps of CC. The black triangle represents the receiver (a MIMO BS), while the colored circles represent the transmitter (UE) locations. The neighborhood relationships—indicated by different colors—in the spatial geometry are preserved and remain in the channel chart.

at the receiver (BS) resides in the radio geometry, since it depends on the radio characteristics of the environment. From the CSI, the receiver extracts features which conform the feature geometry. These features are a representation of the spatial geometry, but reside on a higher dimensional space. To construct a channel chart, therefore, a dimensionality reduction algorithm is applied on the features, with the low-dimensional chart being a representation of the original spatial geometry, in a way such that local geometry is preserved between both geometries. The quality of the representations, that is, the feature and the channel chart geometries can be assessed with the use of quality scores, which measure the performance based on the desired properties.

It is worth mentioning the dimensionality in which the mentioned geometries lie. It is obvious that the spatial geometry can only have dimension $s' = 2$ or 3. CSI resides in a geometry of high dimensionality, and contains complex values with dimension s. This dimension is given by the number of time and/or frequency samples of the signal received at the MIMO BS, which in turn comprises the dimension of the CSI. The dimension of the features, i.e., p, is also relatively high, as it is a transformation of the high dimensional CSI. The channel chart is constructed in a way so that its dimension p' is significantly lower than that of the features. To be able to visualize and better analyze the chart, the typical values for p' are 2 or 3.

For the purpose of extracting meaningful high-dimensional features to be used to construct the chart, numerous approaches have been object of research in the

literature—not necessarily for channel charting, such as based on processing covariance matrix dissimilarities [3, 17]—which capture the large-scale variations of the radio channel—, autoencoders [18], signal subspace approaches [19–21], or on the Multiple Signal Classification (MUSIC) algorithm [21, 22]. It is important to notice that some of the mentioned approached, such as based on AoA, TDoA and MUSIC, intrinsically require a complete knowledge of the structure of the antenna arrays.

3 Problem description

This chapter consists of an extensive description of the methods and tools used throughout the development of this thesis work. This thesis is focused on analyzing and solving, with the means explained in the current chapter, the following problem: Provided with real outdoor CSI measurements and transmitter locations, construct, in an unsupervised manner —except when specified otherwise—, channel charts whose geometry represents that of the true locations. Leverage these charts so that they can be used for localization. That is, estimate the position of transmitters for which their location is unknown, using the CSI associated with them. Evaluate the results with the use of defined performance scores and make the necessary comparisons and evaluations to understand and explain the results. Finally, draw insightful conclusions from said results.

The sections contained in this chapter follow the workflow of CC, depicted in Figure 1, and for each step, a description on what has been done is included, and a brief conclusion if suitable. Firstly, the focus lies on the analysis of the characteristics and properties of the available dataset. Then, the interest is put on how to extract features from the mentioned CSI. Finally, DR algorithms to construct charts are detailed.

3.1 Dataset analysis

This section contains a description of the data used in this thesis, and an analysis of its contents. It is crucial to thoroughly describe and understand the details and particularities of this dataset, in order to be able to evaluate the eventual results with a more profound insight.

As discussed in the introduction, CC is a framework in which a multi-antenna BS can effectively learn the radio geometry of the radio cell, based on uplink CSI collected from transmitters at different locations in its surrounding area. Therefore, the data needed to create charts will primarily consist of wireless CSI.

The dataset utilized in this thesis was originally provided by the IEEE Communications Theory Workshop (CTW) 2020 data competition [23]. It contains a labeled dataset (i.e. CSI with the corresponding 3D ground truth location) and a larger set of unlabeled data for which the true location is unknown. In this thesis most of the work is done using only the labeled dataset.

The CSI measurements were taken in an environment consisting of an outdoor residential area of approximately 0.5 km^2, depicted in Figure 2. It is important to remark that, while this dataset was originally intended to be used to perform user localization, the CSI was not collected with the particular prospect of CC, i.e., there could be different optimizations and parameters which would give more leverage to the construction of radio charts that are not present in the dataset [24].

3.1.1 Measurement parameters

The data are made up of channel frequency responses between a moving transmitter and a receiver which consists of a MIMO 8×8 horizontally polarized patch antenna

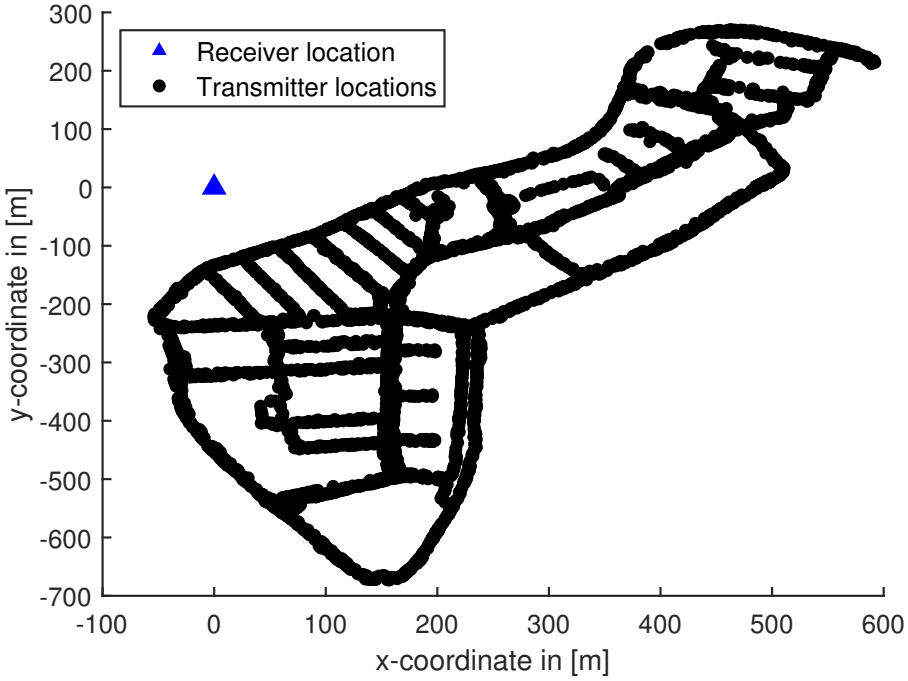


Figure 2: Receiver and transmitter locations. The height dimension is not shown.

array. The uplink Orthogonal Frequency Division Multiplexing (OFDM) pilots used for the channel estimation have a bandwidth of 20 MHz and are made of 1024 subcarriers, of which 100 are used as guard bands. The carrier frequency is set at 1.27 GHz. The CSI was collected using a channel sounder [25], which uses a low-cost flexible architecture capable of measuring CSI in numerous massive MIMO configurations and frequency bands. In particular, the measurements contained in the dataset correspond to the estimation of the frequency response.

The labeled data contain the ground truth coordinates, acquired with a differential Global Positioning System (GPS) device. The location is presented in a XYZ coordinate system, for which the receiver is placed at the origin of the XY-plane. The Z coordinate represents the height. However, and since the measurements were taken in an almost flat surface (compared to the other two dimensions), the height dimension has been ignored in the tasks performed in this thesis.

3.1.2 Dataset description

The data contains a total of 4979 labeled and 36912 unlabeled CSI measurements. Each of those has the following dimensionality:

- $S = 5$ samples of CSI measurements at each location, taken consecutively.
- $F = 924$ subcarriers, after removing those used as guard bands.
- Despite the antenna array containing 64 antennas (8×8), there are 8 antennas which were continuously malfunctioning, so there is only information about

the remaining $A = 56$ antennas. It is important to notice that the position of the malfunctioning antennas in the array is not provided.

Additionally, the Signal-to-Noise-Ratio (SNR) measurement in [dB] at each antenna is also provided, for each of the samples.

For the purpose of generalization, and since the methods described in the following sections do not depend on the dimensionality of this particular dataset, the variables described on Table 1 will be used.

Table 1: Variables defining the dataset dimensionality

Variable	Description
N	Number of transmitter locations.
A	Number of antennas in the array.
F	Number of frequency samples.
S	Number of CSI measurements per location.
When in lowercase, these variables refer to a particular element, as index, i.e., location n for $n = 1, \ldots, N$, antenna a for $a = 1, \ldots, A$, subcarrier f for $f = 1, \ldots F$, and sample s for $s = 1, \ldots, S$.	

To sum up, the CSI data consists of:

- A complex tensor of size $[N, A, F, S]$ containing the estimated channel responses.
- A real tensor of size $[N, A, S]$ with the SNR measured at each antenna.

3.1.3 CSI notation

The notation related to CSI is detailed in Table 2, following the dimensionality notation from Table 1.

3.1.4 Computed power and provided SNR comparison

As explained before, the CSI contains both the channel response matrices and SNR measurements. As SNR is directly proportional to the power, it is important to see if this relationship holds true for the provided SNR and the power calculated from the channel matrices, and understand if they hold exactly the same information or whether there are differences between the two measurements and how they can be used for constructing suitable features which capture the radio characteristics of the channel.

The total received power for sample s at a given location n can easily be calculated from the estimated CSI as:

$$P_{n,s} = \sum_{a=1}^{A} \|\mathbf{h}_{n,a,:,s}\|^2. \tag{1}$$

Table 2: Notation related to CSI

$h_{n,a,f,s} \in \mathbb{C}$	Complex channel response from a particular transmitter location n, corresponding with antenna a, subcarrier f and sample s.
$\mathbf{h}_{n,a,:,s} \in \mathbb{C}^F$	CSI from a particular transmitter location n, antenna a and sample s, for all subcarriers.
The channel vector $\mathbf{h}$ has the dimensions of the position of the index expressed as ':', representing all elements corresponding to that position. For instance, $\mathbf{h}_{n,a,f,:} \in \mathbb{C}^S$ is a vector representing the channel parameters of all samples, corresponding to location n, antenna a and subcarrier f.	
$\mathbf{H}_{n,:,:s} \in \mathbb{C}^{A \times F}$	Channel responses related to location n and sample s, for all subcarriers and antennas.
$\mathbf{R}_{n,s} \in \mathbb{C}^{A \times A}$	Covariance matrix related to location n and sample s. The definition is found in section 3.1.7.
Analogously, when 2 indices are omitted, the CSI matrix is denoted as $\mathbf{H}$, and has the dimensions of the indices expressed as ':'. SNR values use a similar notation, with the exception that the indices are n, a, s, since the values for each subcarrier are not provided, as detailed in section 3.1.2. For instance, $\text{SNR}_{n,a,s}$ denotes the provided SNR value corresponding to location n, antenna a and sample s.	

Since there is more than one sample for every location, it is logical to compare the averages of the power calculated from the CSI and the SNR values. The calculated power for all of the S samples is averaged, $\bar{P}_n = \frac{1}{S}\sum_{s=1}^{S} P_{n,s}$. The SNR corresponding to a location is also averaged, $\overline{\text{SNR}}_n = \frac{1}{S}\sum_{s=1}^{S}\sum_{a=1}^{A} \text{SNR}_{n,a,s}$. Figure 3 illustrates this comparison by observing these averages for all locations. In doing this, it is clearly observable that both variables are strongly correlated, as the correlation coefficient $r = 0.9794$. Therefore, it can be concluded that the inclusion of SNR values in the dataset does not provide any extra information than the one that can be extracted from the CSI measurements alone. Consequently, the features used in CC and localization will be constructed solely based on the CSI values.

3.1.5 Samples and antennas with no received power

A significant fact about the CSI is that there is a significant number of antennas and samples for which the estimated CSI is recorded as a value of zero due the sensitivity of the receiver.

For the labeled data in particular, around 25% of the transmitter locations present at least a CSI sample in which the channel estimation has been reported as 0 for all antennas. This aspect is more present in those locations which are further away from the receiver. However, it is worth noticing that despite the fact that there exist

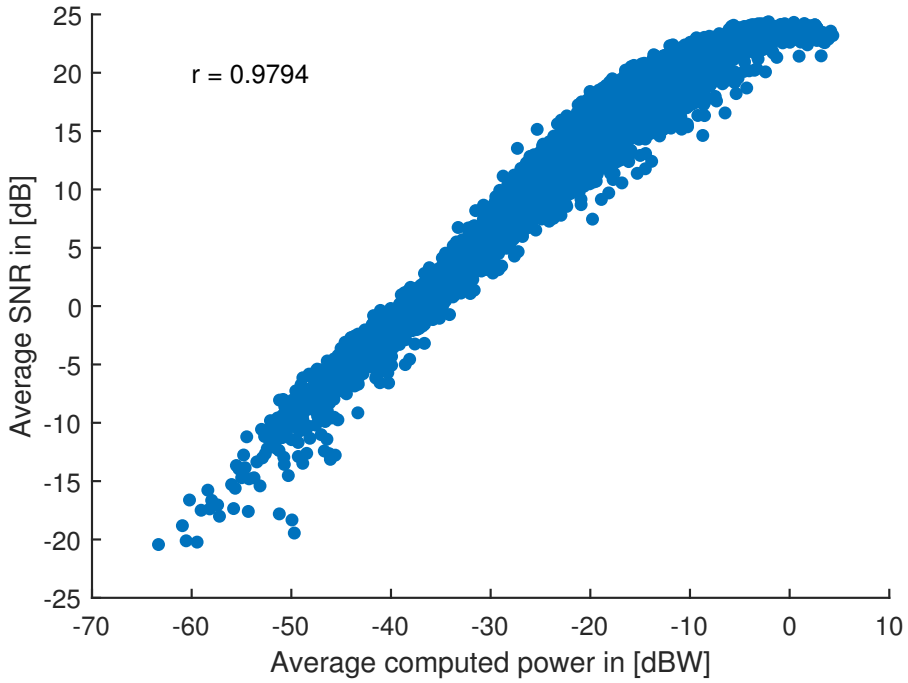


Figure 3: Shows the strong correlation between the averages of computed power and measured SNR. r is the correlation coefficient between the two variables.

samples with no effective CSI, all labeled CSI measurements contain at least one sample in which there are non-zero channel parameters (at least for one antenna). Additionally, it is important to remark that the no-power situation does not occur in the subcarrier domain. That is to say, for those samples with received power, the data contains channel parameters for *all* subcarriers.

This circumstance can be notable on the account of making comparisons between the channels at two different locations, especially if said locations are in a non-Line of Sight (NLoS) situation and contain a small number of non-zero values. For instance, it is hard to make conclusions about the differences between channels at one location with one antenna effectively working and another location with another antenna being the one with non-zero values. Because of this fact, there will be differences in the quality of the information which will be used to construct the features, between the Line of Sight (LoS) and NLoS locations. It can be said that, in general, the locations in LoS will have features which will capture better the underlying radio channel characteristics than those in NLoS.

Clearly, having actual values for those samples and antennas in which the estimation was not possible would allow for more accurate analyses and provide a more insightful overview of the characteristics of the channels at different locations, which would in turn provide better features to solve the CC and localization problems.

3.1.6 Variation of the power over the area

Another important aspect of the analysis of the data is the variation of computed power over the measurements area. As concluded in section 3.1.4, there is no need to analyze both the calculated power from the CSI and the provided SNR values as the computed power and the provided SNR are highly correlated, so only the computed power is analyzed.

Figure 4 shows the average computed power for all locations. For a location n, its corresponding average power is $\bar{P}_n$, as detailed in section 3.1.4. It is observable that, as expected, the locations corresponding to higher power values are majorly situated at the proximity of the receiver, and could be considered to have LoS with it. The locations with the lowest power values are the ones further away from the receiver, i.e. NLoS locations.

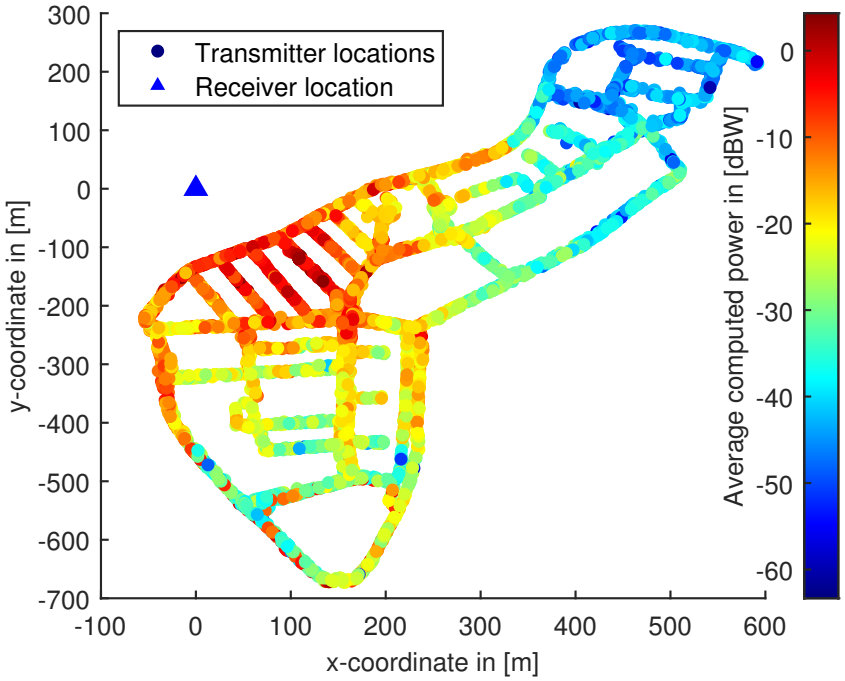


Figure 4: The average computed power as a function of the transmitter location. The blue triangle marks the receiver location. The average received power is higher for those transmitters closer to the receiver (LoS), and lower for those which are far away (NLoS).

3.1.7 Covariance matrices

Covariance matrices are a powerful tool to further understand the meaning of CSI, since they contain useful information that can be exploited for several purposes, including the analysis of wireless channels, feature extraction and localization based on CC. In addition, covariance matrices provide a way of compressing the CSI.

The MIMO recieved signal is described as:

$$\mathbf{y}_{n,:,f,s} = \mathbf{h}_{n,:,f,s}x + \mathbf{w}, \tag{2}$$

where $\mathbf{y}_{n,:,f,s} \in \mathbb{C}^A$ and $x \in \mathbb{C}$ represent the received and transmitted signals, respectively, and $\mathbf{w} \sim CN(0, \sigma^2\mathbf{I}_A)$ denotes the Additive White Gaussian noise (AWGN), with element-wise variance σ^2.

$\mathbf{h}_{n,:,f,s} \in \mathbb{C}^A$ characterizes the MIMO channel between the transmitter and receiver. If the transmitted and received symbols are known, the channel can be estimated, thus obtaining an estimation $\hat{\mathbf{h}}_{n,:,f,s} \approx \mathbf{h}_{n,:,f,s}$. This estimation is outside the scope of this thesis.

The covariance matrix associated to the channel response at a given location n and sample s is computed as

$$\mathbf{R}_{n,s} = \mathbb{E}_f \left[\mathbf{h}_{n,:,f,s}\mathbf{h}^H_{n,:,f,s}\right] \in \mathbb{C}^{A\times A}, \tag{3}$$

where $\mathbb{E}[\cdot]$ denotes expectation, which in this case is with respect to frequency samples, and $(\cdot)^H$ is the Hermitian operation of a matrix or vector.

It is worthwhile to discuss some of the properties of covariance matrices. They are hermitian, i.e., $\mathbf{R}^H_{n,s} = \mathbf{R}_{n,s}$, and positive semi-definite, i.e., $\mathbf{z}\mathbf{R}_{n,s}\mathbf{z} \geq 0 \ \ \forall \ \ \mathbf{z} \in \mathbb{C}^A \neq \mathbf{0}$. They have real non-negative eigenvalues $\lambda_1, \ldots, \lambda_A$, which define a diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_A)$. The eigenvectors $\{\mathbf{v}_i\}_{i=1}^A$ form an orthonormal basis. The eigendecomposition of a covariance matrix can be expressed as $\mathbf{R}_{n,s} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_A]$, and the matrix $\mathbf{V}$ is unitary, i.e., $\mathbf{V}^{-1} = \mathbf{V}^H$.

3.1.8 Variations of the magnitude and phase for different samples and antennas

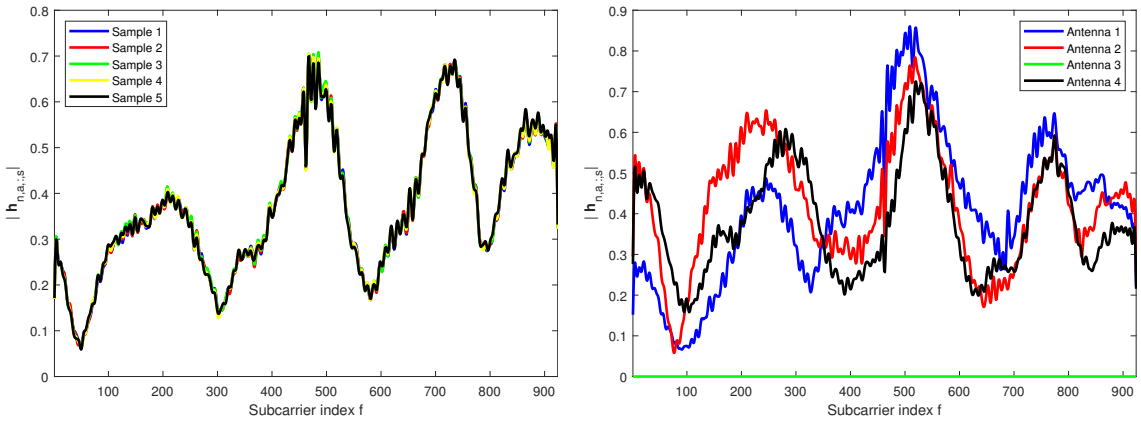
Because of the fact that for each location there is CSI corresponding to different antennas and time samples, an essential step in the analysis of the data is to understand how CSI varies, both in the sample and in the antenna domains. In particular, the interest lies in examining the CSI values (both in magnitude and phase) in order to reveal patterns and behaviors that can affect feature extraction and consequently CC. This analysis focuses on two aspects, which can be done for each transmitter:

- Analyzing the differences in magnitude and phase of the channel corresponding to each of the samples, i.e., $\{\mathbf{h}_{n,\mathfrak{a},:,s}\}_{s=1}^S$, for a given antenna $\mathfrak{a}$.
- Analyzing the differences in magnitude and phase of the channel corresponding to each of the antennas, i.e., $\{\mathbf{h}_{n,a,:,\mathfrak{s}}\}_{a=1}^A$, for a given sample $\mathfrak{s}$.

It is logical to assume that the CSI corresponding to each of the S samples and a particular antenna will be similar, if the assumption that the channel is time-invariant is satisfied. However, this condition can only be true for those samples for which the receiver has been able to measure and record the CSI values. Due to the condition described in 3.1.5, those samples for which the measured channel is 0 cannot be

compared to those for which this condition does not occur. Similarly, the channel observed by each of the A antennas will also be similar. It is important to point out that it is expected that there exists a phase difference between these observations, corresponding to different angles of arrival of the signal at the receiver. However, and since the array locations of the antennas are unknown, it is not trivial to determine the positions of the 8 antennas which are malfunctioning.

Figure 5 depicts the magnitude of the channel response at given locations, for the two described perspectives.



(a) (b)

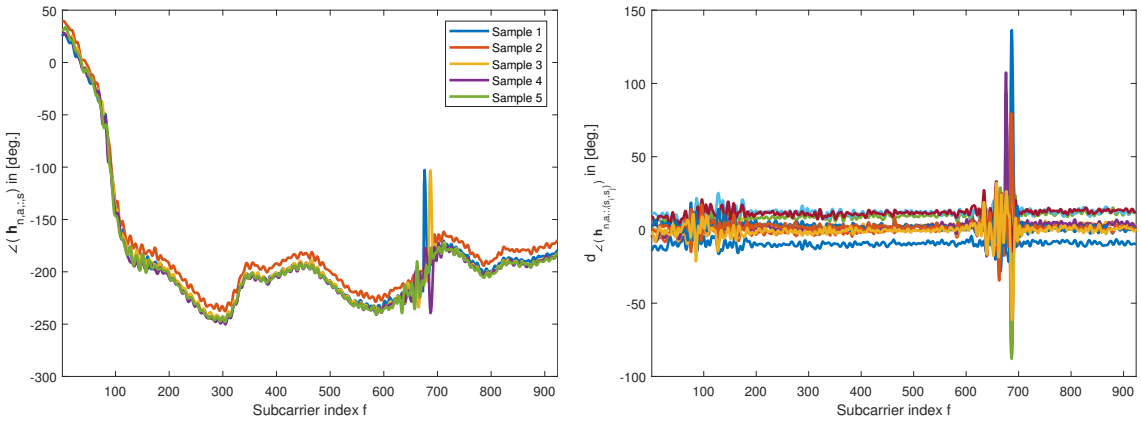
Figure 5: (a) The magnitude of the channel response corresponding to one antenna, for all samples, at a given location. In this case, the difference between samples is negligible. (b) The magnitude of the channel response corresponding to one sample, for a total of 4 antennas, at a given location. In this example, one antenna reported no power. Furthermore, it can be seen that the magnitudes are similar, and mainly affected by a phase shift.

An observation that is important when analyzing the CSI, especially when it can be assumed that there is a non-neglectable amount of noise in the data —the effect of having no-power situations is an example— is to look at the phase information. There might exist phase offset errors that would make feature extraction more difficult, and thus phase correction algorithms used to correct such errors might be needed, as discussed in [26].

One way of visualizing the behavior of the phase differences is to analyze if such difference depends on the frequency. For a particular antenna, the difference in phase between the different samples should be constant.

For this, the phase differences are computed, $\mathbf{d}_{\angle_{n,a,:,(s_i,s_j)}} = \angle(\mathbf{h}_{n,a,:,s_i}) - \angle(\mathbf{h}_{n,a,:,s_j}) \in \mathbb{R}^F$, for those $i, j = 1, \ldots, S, i \neq j$ where the phase is defined. This gives a maximum of $\binom{S}{2}$ phase differences for each location.

Figure 6 illustrates the above mentioned concept, for a particular location and antenna. It can be observed that the phase difference between antennas does not have a dependency on the frequency. This fact extends to all locations, in general. Therefore, the quality of the features will not be affected by phase offsets, and there



(a) (b)

Figure 6: Shows (a); the phase of the channel response corresponding to one antenna, for all samples, at a given location. (b); the 10 pairwise phase differences from (a). In this example, the phase difference between samples can be considered constant, i.e., it does not depend on the frequency. The presence of noise is clearly visible, but in this case it affects a small part of the spectrum.

is no need for applying phase correction as a step prior to feature extraction.

3.1.9 Frequency selectivity of the channel

The study of the frequency selectivity of the radio channel can help understand the fading conditions in the band of operation. Observing flat-fading channels, in which the frequency components (each subcarrier) experience a similar amount of fading, would make sure that the signal can be properly received, thus allowing for better channel estimations and, therefore, radio features that represent the physical geometry in a more accurate manner. In order to measure the radio frequency selectivity at a given location, the dissimilarity between the estimated channels at consecutive subcarriers is defined as:

$$d^2_{n,s,f,f+1} = 1 - \left| \tilde{\mathbf{h}}^H_{n,:,f,s,} \tilde{\mathbf{h}}_{n,:,f+1,s} \right|^2, f = 0 \dots F-1, \tag{4}$$

where $\tilde{\mathbf{h}}_{n,:,f,s} = \frac{\mathbf{h}_{n,:,f,s,}}{\left\|\mathbf{h}_{n,:,f,s}\right\|} \in \mathbb{C}^A$ is the normalized CSI, i.e., $\left\|\tilde{\mathbf{h}}_{n,:,f,s}\right\| = 1$.

The value of $d_{n,s,f,f+1}$, for a particular sample s, will be low, i.e., $d_{n,s,f,f+1} \ll 1$, if the channels corresponding to consecutive subcarriers f and $f+1$ are similar, i.e. the channel response can be considered to be flat around those frequencies.

In order to visualize the frequency selectivity, it can be done by observing the histogram of all $F-1$ values, for each of the samples, at each particular location. Figure 7 illustrates an example related to a particular transmitter location. The $d_{n,s,f,f+1}$ values can be computed for all locations in order to obtain a more general picture of the radio characteristics of the area.

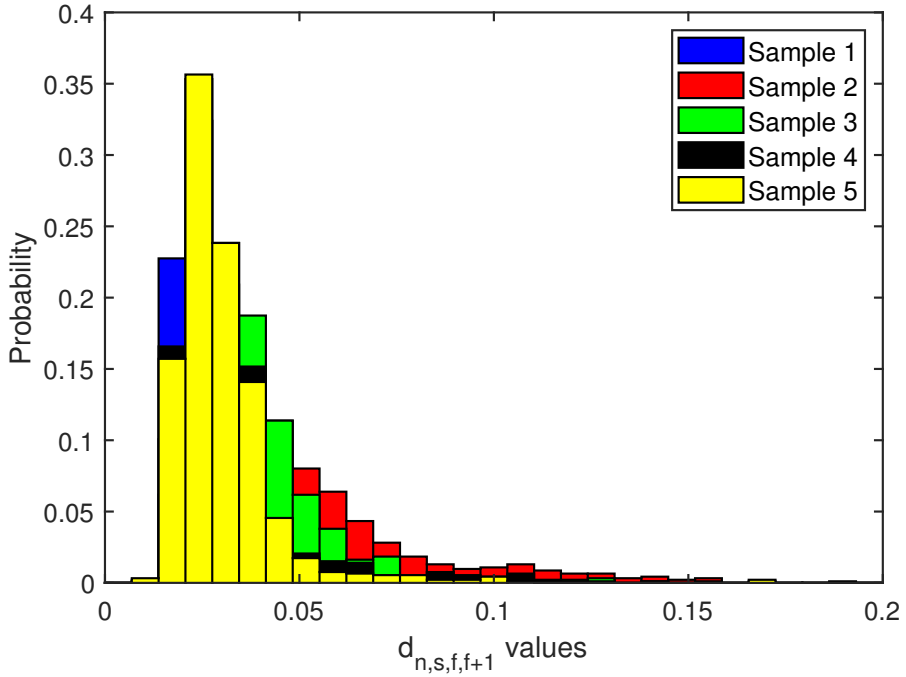


Figure 7: An example of frequency selectivity visualization for a particular location. In this case, the channel can be considered to be non-frequency selective, due to the relatively low values of d for all 5 samples.

3.2 Quality scores for features and charts

There is a need to introduce certain measures to determine the convenience of the used channel features, as well as to assess the quality of the constructed channel charts. In particular, the focus is on how well the geometry of the channel chart can represent the spatial geometry of the true locations. High quality charts will be those who preserve the neighborhood relations as much as possible, so that points which are nearby on the chart represent nearby transmitter locations.

There exist metrics which can be used to calculate the quality of the channel features and the learned charts [27–30]. In this thesis three of them are used. The first two, namely trustworthiness (TW) and continuity (CT) [27], measure how well neighborhood is preserved. A third metric, Kruskal's Stress [28], measures how much the chart distances are distorted when compared to the distances between the transmitter locations. In the scheme of CC it is a measure of how well distances in the chart can represent the spatial geometry.

In order to describe these metrics, two sets of N points are defined. Namely, the original space $\mathcal{U}$ and the representation space $\mathcal{V}$. In the framework of CC, let $\mathcal{U}$ be the set corresponding to the spatial geometry, and $\mathcal{V}$ a representation of the said original space, i.e., the true locations, in a way such that an element $v_n \in \mathcal{V}$ represents the element $u_n \in \mathcal{U}, n = 1 \dots N$. $\mathcal{V}$ can either contain the points on the channel chart (for measuring its quality), or the feature geometry (for measuring the quality of the features used to construct the chart). Table 3 contains the notation and terms for defining the quality scores that are used in this thesis.

Table 3: Notation related to quality scores

$d_{\mathcal{U}}$	A distance defined between pairs of elements in $\mathcal{U}$. In the framework of CC, $d_{\mathcal{U}}$ is the pairwise true distance between two locations.
$d_{\mathcal{V}}$	A distance defined between pairs of elements in $\mathcal{V}$. In the framework of CC, $d_{\mathcal{V}}$ is the distance between points in the channel chart, or a distance/dissimilarity between CSI features, which are defined in section 3.3.
K-neighborhood	Set of the K closest neighbors of a point $u_n \in \mathcal{U}$, according to a distance $d_{\mathcal{U}}$, and analogously for a point $v_n \in \mathcal{V}$.
$\mathcal{V}_K(u_n)$	Set containing those elements in the K-neighborhood of u_n whose representations are not in the K-neighborhood of v_n.
$\mathcal{U}_K(v_n)$	Set containing those elements in the K-neighborhood of v_n which are not in the K-neighborhood of the corresponding point u_n in the original space.
$r(u_n, u_m)$	Rank of point u_m in the K-neighborhood of u_n when the neighbors are ordered according to the distance $d_{\mathcal{U}}$. For instance, $r(u_n, u_m) = l$ means that point u_m is the l-th nearest neighbor of u_n.
$\hat{r}(v_n, v_m)$	Rank of point v_m in the K-neighborhood of v_n when the neighbors are ordered according to the distance $d_{\mathcal{V}}$.

3.2.1 Continuity (CT)

Continuity is a measure of how much the neighbors of points in the original space are preserved in the representation space, i.e., it measures how much neighbors of a point in the original space are also neighbors in the representation space. The point-wise continuity for K neighbors between a data point u_n and its representation v_n is [27]:

$$\mathrm{CT}_n(K) = 1 - \frac{2}{K(2N - 3K - 1)} \sum_{m \in \mathcal{V}_K(u_n)} (\hat{r}(v_n, v_m) - K). \tag{5}$$

The continuity between the sets $\mathcal{U}$ and $\mathcal{V}$ is simply the average of the point-wise continuities, $\mathrm{CT}(K) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{CT}_n(K)$. The point-wise continuity and, by definition, the average continuity can take values between 0 and 1. An average continuity value of 1 indicates that all neighborhood relationships which are present in the original set *continue* to exist in the representation set.

3.2.2 Trustworthiness (TW)

Trustworthiness is a measure of how much the local geometry of the representation space can be *trusted* to represent the geometry of the original space, i.e., how many of the neighbors in the representation space are also neighbors in the original space.

The point-wise trustworthiness for K neighbors between points u_n and v_n is [27]:

$$\mathrm{TW}_n(K) = 1 - \frac{2}{K(2N - 3K - 1)} \sum_{m \in \mathcal{U}_K(v_n)} (r(u_n, u_m) - K). \tag{6}$$

The trustworthiness between the original and the representation sets is the average of the point-wise trustworthiness values, $\mathrm{TW}(K) = \frac{1}{N} \sum_{n=1}^{N} \mathrm{TW}_n(K)$. The point-wise trustworthiness and, by definition, the average trustworthiness can take values between 0 and 1. A high value indicates that the neighbor relationships of points in the representation space $\mathcal{V}$ can be *trusted* to exist in the original space.

3.2.3 Kruskal's Stress (KS)

Kruskal's Stress is a measure of *goodness of fit*, i.e. it measures how well the representation space can represent the original space. The stress between the two sets is defined as [28]:

$$\mathrm{KS}(\mathcal{U}, \mathcal{V}) = \sqrt{\frac{\sum_{n<m} \left(d_{\mathcal{U}}(u_n, u_m) - d_{\mathcal{V}}(v_n, v_m) \right)^2}{\sum_{n<m} d_{\mathcal{U}}(u_n, u_m)^2}}. \tag{7}$$

This metric has values between 0 and 1, where a stress of 0 indicates a perfect monotone relationship between the distances (or dissimilarities) of the original and representation sets $\mathcal{U}$ and $\mathcal{V}$. This metric is invariant under rigid transformations such as rotation, translation and reflections. The normalization term in the denominator ensures that the metric is also invariant under a uniform scaling of the sets.

3.3 Dissimilarity measures and feature construction

In order to proceed with CC and localization, there need to be found features taken from the CSI which represent the underlying unknown geometry, i.e., the true location in the most accurate way possible. That is, construct a representation space from the CSI features in which neighborhood is preserved and where the distances are distorted the least, compared to the original space. As previously mentioned, the original space, in this context, is made up of the true locations.

Related to the score values, it is equivalent to say that features that accomplish those two requirements (neighborhood preserving and low distance distortion) will have better quality scores than those features whose geometry fails to preserve the spatial geometry of the transmitter locations.

3.3.1 Covariance matrix distances

A natural way of considering the creation of features is to start from the covariance matrices defined at each location, $\mathbf{R}_n, n = 1, \ldots, N$.

There exist numerous metric functions which define distances (dissimilarities) between matrices. By utilizing these distance functions it is possible to create features that might be related to the distances between true locations. Distance functions

which lead to features whose geometry is an accurate representation of the spatial geometry of the true locations (according to the quality scores) will be preferable. A distance d between two matrices is simply

$$d : \mathbb{C}^{A\times A} \times \mathbb{C}^{A\times A} \longrightarrow \mathbb{R}. \tag{8}$$

The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{C}^{M\times N}$, which is used used in some of the distance functions, is a matrix norm defined as:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N} |[\mathbf{A}]_{ij}|^2} = \sqrt{\mathrm{Tr}(\mathbf{A}\mathbf{A}^H)}, \tag{9}$$

where $[\mathbf{A}]_{ij}$ denotes the element of $\mathbf{A}$ at the ith row and jth column. The Frobenius norm has all the properties of a matrix norm.

In this thesis, several of already defined distances are utilized and their performances are evaluated. The following list contains a description of each.

- **Frobenius distance**

  The Frobenius distance d_F between two matrices $\mathbf{R}_n$ and $\mathbf{R}_m$ can be defined in terms of the Frobenius norm,

  $$d_F\left(\mathbf{R}_n, \mathbf{R}_m\right) = \left\|\mathbf{R}_n - \mathbf{R}_m\right\|_F. \tag{10}$$

- **Correlation Matrix Distance (CMD)**

  CMD between two covariance matrices is defined as [31]:

  $$d_{\mathrm{CMD}}(\mathbf{R}_n, \mathbf{R}_m) = 1 - \frac{\mathrm{Tr}(\mathbf{R}_n\mathbf{R}_m)}{\|\mathbf{R}_n\|_F\|\mathbf{R}_m\|_F}, \tag{11}$$

  where $\mathrm{Tr}(\cdot)$ denotes the trace operation.

- **Chordal Distance**

  Let $\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \in \mathbb{C}^{N\times N}$ be the eigendecomposition of the covariance matrix $\mathbf{R}$. $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_N]$ is the matrix containing the eigenvectors corresponding to the eigenvalues ordered in descending order. Let $\mathbf{V}_k = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$ and $\mathbf{U} = \mathbf{V}_k\mathbf{V}_k^H$. The chordal distance between two covariance matrices is defined as [32]:

  $$d_{\mathrm{chord}}(\mathbf{R}_n, \mathbf{R}_m) = \|\mathbf{U}_n - \mathbf{U}_m\|_F. \tag{12}$$

- ***X* Distance**

  The X distance is derived from [33] and is defined by first introducing a vector, $\mathbf{x}_\mathbf{R}$, associated to covariance matrix $\mathbf{R}$, as:

  $$\mathbf{x}_\mathbf{R} = \frac{\left[\Re\Big(\mathrm{vec}\left(\mathbf{R}_{\mathrm{UTD}}\right)\Big)^T, \Im\Big(\mathrm{vec}\left(\mathbf{R}_{\mathrm{UT}}\right)\Big)^T\right]^T}{\mathrm{Tr}(\mathbf{R})}, \tag{13}$$

where the vec$(\cdot)$ operation takes the elements in a matrix and stacks them in a column vector, $\mathbf{R}_{\text{UTD}}$ is the upper triangular side of $\mathbf{R}$, with its diagonal, and $\mathbf{R}_{\text{UT}}$ is its upper triangular part. $\Re$ and $\Im$ are the real and imaginary parts, respectively.

The X distance between two covariance matrices is given by:

$$d_X(\mathbf{R}_n, \mathbf{R}_m) = \|\mathbf{x}_{\mathbf{R}_n} - \mathbf{x}_{\mathbf{R}_m}\|_2 . \tag{14}$$

- $\mathbb{C}\{\cdot\}$ **Distance**

  This distance is proposed in [3] and given by:

$$d_{\mathbb{C}\{\cdot\}}(\mathbf{R}_n, \mathbf{R}_m) = d_F(\hat{\mathbf{R}}_n, \hat{\mathbf{R}}_m), \quad \text{where } \hat{\mathbf{R}} = \frac{A^{\beta-1}}{\|\mathbf{R}\|_F^{\beta}}\mathbf{R}, \tag{15}$$

  where A is the number of antennas in the array and $\beta = 1 - \frac{1}{2\alpha}$, with α representing the path loss exponent, so that the CSI is inversely proportional to the true distance between the transmitter and receiver: $h_{n,a,f,s} \propto d_{\text{true}}^{-\alpha}$, for $\alpha \in \mathbb{R}_+^*$. In this thesis, for the particular dataset in question, the value utilized is $\alpha = 8$, as it gives the best results in terms of feature quality.

3.3.2 Signal subspace based dissimilarities

A different approach in the quest for finding features is the use of dissimilarities based on signal subspaces [19–21].

At every location, the signal subspace is defined as the subspace of smallest dimension onto which a certain percentage of the received energy projects. Firstly, the assumption that the signals and noise are uncorrelated needs to be made.

The covariance matrix can be expressed as a combination of a signal and additive noise, which will be considered as existing in two orthogonal subspaces, namely the signal and the noise subspaces.

$$\mathbf{R}_n = \mathbb{E}_{f,s}\left[\mathbf{y}_{n,:f,s}\mathbf{y}_{n,:f,s}^H\right] = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^H + \sigma^2\mathbf{I}, \tag{16}$$

where $\mathbf{A}$ contains direction of arrival information, and $\mathbf{\Sigma}$ represents the power of the multi-path components of the received signal. The normalized version of the covariance matrix is defined as

$$\bar{\mathbf{R}}_n = \frac{\mathbf{R}_n}{\text{Tr}(\mathbf{R}_n)}, \tag{17}$$

and after the trace normalization, $\text{Tr}(\bar{\mathbf{R}}_n) = 1$.

The signal subspace is defined as the subspace spanned by the eigenvectors of $\bar{\mathbf{R}}_n$ corresponding to the Q largest eigenvalues.

The dimension of the signal subspace Q is defined in a way such that the said subspace captures a certain fraction $\rho \in [0, 1]$ of the total energy, i.e. $\sum_{k=1}^{A} \lambda_k = 1$. In particular, and for the features used in this thesis, the signal subspaces are constructed

with a dimension which is able to capture at least 90% of the energy, i.e., $\rho = 0.9$. Thus,

$$Q = \arg\min_{Q} \left(\sum_{k=1}^{Q} \lambda_k \geq \rho \mid Q = 1, \ldots, A \right). \tag{18}$$

The projection matrix $\mathbf{P}_n \in \mathbb{C}^{A \times A}$ is defined as the projection matrix that projects the received signal onto the signal subspace,

$$\mathbf{P}_n = \mathbf{V}_Q (\mathbf{V}_Q^H \mathbf{V}_Q)^{-1} \mathbf{V}_Q^H = \mathbf{V}_Q \mathbf{V}_Q^H. \tag{19}$$

where $\mathbf{V}_Q \in \mathbb{C}^{A \times Q} = [\mathbf{v}_1, \ldots, \mathbf{v}_Q]$, and the property that the eigenvectors form an orthonormal basis is used, $\mathbf{V}_Q^H \mathbf{V}_Q = \mathbf{I}_A$.

It is worth noticing the following result:

$$\mathrm{Tr}(\mathbf{P}_n \bar{\mathbf{R}}_n) = \sum_{k=1}^{Q} \lambda_k = \mu \geq \rho, \tag{20}$$

which indicates that the trace of the resulting projection is a measure of what fraction of the total energy, μ, is in the signal subspace.

Having come to that result, it is natural to define a similarity value based on the projection matrices. For two locations m and n, let

$$f_{mn} = \mathrm{Tr}(\mathbf{P}_n \bar{\mathbf{R}}_m) \in [0, 1], \tag{21}$$

which measures the fraction of energy at location m is present in the signal subspace corresponding to location n. Note that $f_{mn} \neq f_{nm}$.

Under the assumption that the channels corresponding to nearby locations have a greater similarity than those which correspond to distant locations, it is safe to assume that the same behavior will be observed with the signal subspaces. Therefore, it is expected that the signal subspaces are able to capture a greater fraction of the energy of close locations than of far away ones.

As the interest is in finding dissimilarities that can represent the distances between the true locations, a simple symmetric dissimilarity between locations m and n can be defined as

$$\tilde{f}_{mn} = \tilde{f}_{mn} = 1 - \frac{f_{mn} + f_{nm}}{2}. \tag{22}$$

If the previous assumption is correct, it can therefore be expected that $f_{mn} \approx f_{nm}$, so it is sensible to take the average of both as situations where a similar value of $\tilde{f}_{mn}$ happens for two significantly different pairs of values $\{f_{mn}, f_{nm}\}$ are highly unlikely. For instance, $\tilde{f}_{mn} = 0.5$ for pairs $\{1, 0\}$ and $\{0.5, 0.5\}$, but the latter situation is more likely than the former.

3.3.3 Relationship between dissimilarities

It is worthwhile to comment about the possible relationship between the previously introduced dissimilarities, in order to better understand their mathematical meaning

and ensure they have the desired properties and are reasonable from a mathematical point of view.

The Frobenius norm is directly related to the eigendecomposition of covariance matrices. As previously stated, $\|\mathbf{R}\|_F^2 = \mathrm{Tr}(\mathbf{R}\mathbf{R}^H)$.

Let $\mathbf{R} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathbf{H}}$ be the eigendecomposition of $\mathbf{R}$. Following the result from equation (9), the relationship is clear:

$$\|\mathbf{R}\|_F^2 = \mathrm{Tr}(\mathbf{R}\mathbf{R}^H) = \mathrm{Tr}(\mathbf{R}^2) = \mathrm{Tr}(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H) = \mathrm{Tr}(\mathbf{\Lambda}^2) = \sum_{k=1}^{A} \lambda_k^2. \tag{23}$$

An aspect worth commenting is the relationship between the Frobenius norm and the vector norm. Let $\mathbf{r} = \mathrm{vec}(\mathbf{R})$ be a vector containing all entries of a covariance matrix $\mathbf{R} \in \mathbb{C}^{A\times A}$. The Frobenius norm of $\mathbf{R}$ is the Euclidean norm of $\mathbf{r}$, $\|\mathbf{R}\|_F = \|\mathbf{r}\|$. Moreover, the Frobenius norm of a covariance matrix $\mathbf{R}$ can be thought as

$$\|\mathbf{R}\|_F = \left\| \left[\mathrm{diag}(\mathbf{R})^T, \sqrt{2}\Re\Big(\mathrm{vec}(\mathbf{R}_{\mathrm{UT}})\Big)^T, \sqrt{2}\Im\Big(\mathrm{vec}(\mathbf{R}_{\mathrm{UT}})\Big)^T \right]^T \right\|. \tag{24}$$

It is crucial to discuss the importance of utilizing a normalization of the covariance matrices before feature extraction, in order to scale the CSI measurements. The relative scales of CSI vary greatly, mainly due to path loss. It is sensible to remove such a variability in comparing radio channels between distinct locations, since the total received power can have significantly different values. In other words, raw (unnormalized) CSI is a poor representation of the spatial geometry [3].

In order to illustrate why normalization is an important step for obtaining features that better represent the underlying spatial geometry of the transmitters, let A_1 and A_2 be two transmitters located near the receiver, while B_1 and B_2 are located further away. Let the true relative distances inside each set be the same, i.e. $d(A_1, A_2) = d(B_1, B_2)$. Due to path loss, the received power corresponding to set B is smaller compared to that of A, and its CSI measurements have smaller values. In this situation, a dissimilarity which does not contemplate any sort of scaling would determine that $d(A_1, A_2) > d(B_1, B_2)$, due to the fact that the variation of CSI is smaller in B than it is in A, as a consequence of path loss.

Hence, there is a need to compensate for this phenomenon and scale the features in a way such that dissimilarities between the channels at two locations are independent from the received power, thus allowing for a feature geometry which is a better representation of the true transmitter locations.

Two natural ways of normalizing covariance matrices are used in the dissimilarity measures previously described. The CMD and $\mathbb{C}\{\cdot\}$ utilize the Frobenius-normalization, while the signal subspace dissimilarity utilizes the trace normalization. Therefore, the following normalizations can be defined:

$$\tilde{\mathbf{R}} = \frac{\mathbf{R}}{\|\mathbf{R}\|_F} = \frac{\mathbf{R}}{\sqrt{\mathrm{Tr}(\mathbf{R}^2)}} = \frac{\mathbf{R}}{\sqrt{\mathrm{Tr}(\mathbf{\Lambda}^2)}}, \tag{25}$$

and following the trace normalization introduced in (17),

$$\bar{\mathbf{R}} = \frac{\mathbf{R}}{\mathrm{Tr}(\mathbf{R})} = \frac{\mathbf{R}}{\left\|\sqrt{\mathbf{R}}\right\|_F^2} = \frac{\mathbf{R}}{\mathrm{Tr}(\mathbf{\Lambda}^2)}. \tag{26}$$

The CMD can therefore be defined by

$$d_{\mathrm{CMD}}(\mathbf{R}_n, \mathbf{R}_m) = 1 - \mathrm{Tr}(\tilde{\mathbf{R}}_n \tilde{\mathbf{R}}_m). \tag{27}$$

Using the Frobenius normalization, the following result is reached:

$$d_F^2\left(\tilde{\mathbf{R}}_n, \tilde{\mathbf{R}}_m\right) = \left\|\tilde{\mathbf{R}_n} - \tilde{\mathbf{R}}_m\right\|_F = \mathrm{Tr}\left[\left(\tilde{\mathbf{R}}_n - \tilde{\mathbf{R}}_m\right)^2\right] = \mathrm{Tr}\left(\tilde{\mathbf{R}}_n^2 + \tilde{\mathbf{R}}_n^2 - 2\tilde{\mathbf{R}}_n\tilde{\mathbf{R}}_m\right). \tag{28}$$

By using the property from (23), $\mathrm{Tr}\left(\tilde{\mathbf{R}}^2\right) = \left\|\tilde{\mathbf{R}}\right\|_F^2 = 1$, then

$$d_F^2(\tilde{\mathbf{R}}_n, \tilde{\mathbf{R}}_m) = 2 - 2\mathrm{Tr}(\tilde{\mathbf{R}}_n \tilde{\mathbf{R}}_m), \tag{29}$$

and combining this result with (27),

$$d_F^2(\tilde{\mathbf{R}}_n, \tilde{\mathbf{R}}_m) = 2d_{\mathrm{CMD}}(\mathbf{R}_n, \mathbf{R}_m). \tag{30}$$

In the $\mathbb{C}\{\cdot\}$ distance, there is also a normalization in use. This normalization is again related to the Frobenius norm of the covariance matrix. It is worth noting that the $\mathbb{C}\{\cdot\}$ normalization equals the Frobenius norm normalization for $\alpha \to \infty$, with $\beta = 1 - \frac{1}{2\alpha}$, $\lim_{\alpha\to\infty} \beta = 1$.

$$\lim_{\alpha\to\infty} \hat{\mathbf{R}} = \lim_{\alpha\to\infty} \frac{A^{\beta-1}}{\|\mathbf{H}\|_F^{\beta}}\mathbf{R} = \frac{A^0}{\|\mathbf{H}\|_F^1}\mathbf{R} = \tilde{\mathbf{R}}. \tag{31}$$

Thus, $d_{\mathbb{C}\{\cdot\}}(\mathbf{R}_n, \mathbf{R}_m) \to \mathrm{CMD}(\mathbf{R}_n, \mathbf{R}_m)$ when $\alpha \to \infty$. In this case, the relative scaling of the matrices due to path loss vanishes, as happens in the CMD.

The value of α is an indication of how much this path loss related scaling is destroyed, thus allowing for a more flexible normalization whose extreme cases range from no normalization at all, when $\alpha \to 0$, since $\lim_{\alpha\to 0} \hat{\mathbf{R}} = \mathbf{R}$, to a normalization of the Frobenius norm, $\hat{\mathbf{R}} = \tilde{\mathbf{R}}$ for $\alpha \to \infty$.

In regards to the X distance, there is a clear relationship with the Frobenius distance. As the X distance discards one triangular part of the covariance matrix, it can be redefined as

$$d_X(\mathbf{R}_n, \mathbf{R}_m) = d_F\left(\bar{\mathbf{R}}_{n_{\mathrm{UTD}}}, \bar{\mathbf{R}}_{m_{\mathrm{UTD}}}\right) = \sqrt{\frac{1}{2}d_F^2\left(\bar{\mathbf{R}}_{n_{\mathrm{OD}}}, \bar{\mathbf{R}}_{m_{\mathrm{OD}}}\right) + d_F^2\left(\bar{\mathbf{R}}_{n_{\mathrm{D}}}, \bar{\mathbf{R}}_{m_{\mathrm{D}}}\right)}, \tag{32}$$

where $\bar{\mathbf{R}}_{\mathrm{OD}}$ is the off-diagonal part of $\bar{\mathbf{R}}$, and $\bar{\mathbf{R}}_{\mathrm{D}}$ its diagonal part.

From the X distance a norm can be defined, namely the X-norm, given by:

$$\|\mathbf{R}\|_X^2 = \frac{1}{2}\left\|\bar{\mathbf{R}}_{\mathrm{OD}}\right\|_F^2 + \left\|\bar{\mathbf{R}}_{\mathrm{D}}\right\|_F^2, \tag{33}$$

which can be used also as a different normalization.

It can be easily seen that the chordal distance is closely related to the signal subspace based dissimilarity. In the chordal distance, k eigenvectors are used to construct $\mathbf{U}$, whereas the projection matrix which projects the signal onto the signal subspace is constructed with Q eigenvectors depending on the portion of energy captured in the said subspace. It can be observed that $\mathbf{U}$ is also a projection matrix whose associated subspace is of dimension k. Hence, the chordal distance measures dissimilarities between projection matrices, with the particularity that they all project onto subspaces of the same dimension, contrary to what happens in the dissimilarity based on signal subspaces. It is worth following the result from equation (28) in order to further understand how the chordal distance is a dissimilarity between projection matrices.

$$d^2_{\text{chordal}}(\mathbf{R}_n, \mathbf{R}_m) = d^2_F(\mathbf{U}_n, \mathbf{U}_m) = \text{Tr}\left(\mathbf{U}_n^2 + \mathbf{U}_m^2 - 2\mathbf{U}_n\mathbf{U}_m\right). \tag{34}$$

Note that the projection matrix $\mathbf{U}$ is idempotent, i.e. $\mathbf{U}^2 = \mathbf{U}$, and its trace is k. Hence,

$$d^2_{\text{chordal}}(\mathbf{R}_n, \mathbf{R}_m) = 2\left(k - \text{Tr}\left(\mathbf{U}_n\mathbf{U}_m\right)\right) = 2k\left(1 - \text{Tr}\left(\tilde{\mathbf{U}}_n\tilde{\mathbf{U}}_m\right)\right) = 2k\, d_{\text{CMD}}\left(\mathbf{U}_n, \mathbf{U}_m\right). \tag{35}$$

As it can be seen from the equations, the upper bound of the chordal distance is given by $\sqrt{2k}$.

In theory, it could be possible to define a chordal distance in which the matrices $\mathbf{U}$ were the projection matrices $\mathbf{P}$ used in the signal subspace based similarity, where the dimension of the related subspace depends on the distribution of energy among the eigenvalues of the covariance matrices.

More generically, most matrix dissimilarities can be considered to be a form of measuring how the power is distributed between the eigenvalues. By transforming the features, e.g., operations such as normalization, the eigendecomposition of a transformed covariance matrix $\mathbf{R}'$ takes the form

$$\mathbf{R}' = \mathbf{V}\mathbf{\Lambda}'\mathbf{V}^H, \tag{36}$$

where a dissimilarity d_{diss} can simply be expressed in terms of a power $\gamma > 0$ of the Frobenius distance, $d_{\text{diss}}\left(\mathbf{R}_n, \mathbf{R}_m\right) = d^{\gamma}_F\left(\mathbf{R}'_n, \mathbf{R}'_m\right)$

For example, the CMD could be expressed by defining a linear transformation $\mathbf{T}$ on the eigenvalues (in this case a scaling), making $\mathbf{R}' = \mathbf{V}\left(\mathbf{T}\mathbf{\Lambda}\right)\mathbf{V}^H$, with $\mathbf{T} = \frac{1}{2\sqrt{\text{Tr}(\mathbf{\Lambda}^2)}}\mathbf{I}$, and then use the square Frobenius distance ($\gamma = 2$).

3.3.4 Relationship between dissimilarity and physical distance

After having defined several dissimilarities between covariance matrices, it is sensible to evaluate how well they represent the true distances between transmitters. A natural way of visualizing such a relationship is by looking at the correlation between the two. Moreover, it is also worthy to compare the possible differences in correlation for LoS and NLoS locations. It is expected that, in general, a dissimilarity between

two points which are the same distance apart is higher at NLoS locations than at LoS locations, due to the effect of poor channel estimations due to low received power, discussed in 3.1.5. Additionally, it is expected that the signals received from two NLoS transmitters travel in paths that are more dissimilar, i.e. their multipath components have more variability, than the signals related to LoS locations. This fact prevents similarities to be discovered between CSI from closeby NLoS locations, in comparison to that of LoS locations.

This phenomenon might introduce unwanted behavior when comparing dissimilarities, since the MIMO channel estimations in some of the NLoS locations do not contain enough information about the true channel coefficients, and hence a dissimilarity between two poorly estimated channels is highly unreliable, compared to using better estimations like those at LoS locations. If the assumption that the variation of the statistical properties of the MIMO channel is slow across space is made—due to the locations of scatterers in the channel; it is expected that a dissimilarity will have a strong correlation with the true distance, for the cases where the distance is shorter than the distance in which the channels are highly uncorrelated. In other words, if we assume three locations p, q, r with a relative large distance separating them, the correlation between dissimilarity and distance is lost, i.e. $d_{\text{true}}(p,q) \gg d_{\text{true}}(p,r) \not\Rightarrow d_{\text{diss}}(p,q) \gg d_{\text{diss}}(p,r)$ if $d_{\text{true}}(p,q)$ is such that the channels at p and q are uncorrelated and, thus, have a high dissimilarity, comparable to that between the channels at p and r.

As a consequence, a dissimilarity can represent the true distances only when they are relatively low. Beyond that, there is small to no correlation found, and the dissimilarity is a poor representation of true distances.

Note that this fact does not hinder the usefulness of utilizing dissimilarities for constructing features, and it is a behavior which truly describes the radio geometry. Furthermore, neighborhood preservation is not affected. If a dissimilarity is able to represent the short—neighboring—distances accurately at all locations, then the feature geometry will conserve the neighborhood relationships of the spatial geometry, if a suitable DR technique is used, i.e., that preserves short distances.

In Figure 8 the property that dissimilarity is related to distance can be observed. For the smaller distances, it can be seen that there is a linear correlation between both distances. It is also patent that, depending on the location, this correlation will be more or less clear. Additionally, it can be observed that undesired properties will occur, as there might be locations whose channels appear similar to others measured at distant locations.

Figure 9 illustrates the relationship between true distance and CMD from the point of view of the neighborhoods of two transmitters, with one being in LoS with the receiver, and the other in NLoS. For these relatively lower distances, that is, the regime where the dissimilarity depends on the true distance, it is visible that for LoS locations this dependence is significantly stronger than for NLoS locations, where this dependency is weaker and, in some cases, non-existing.

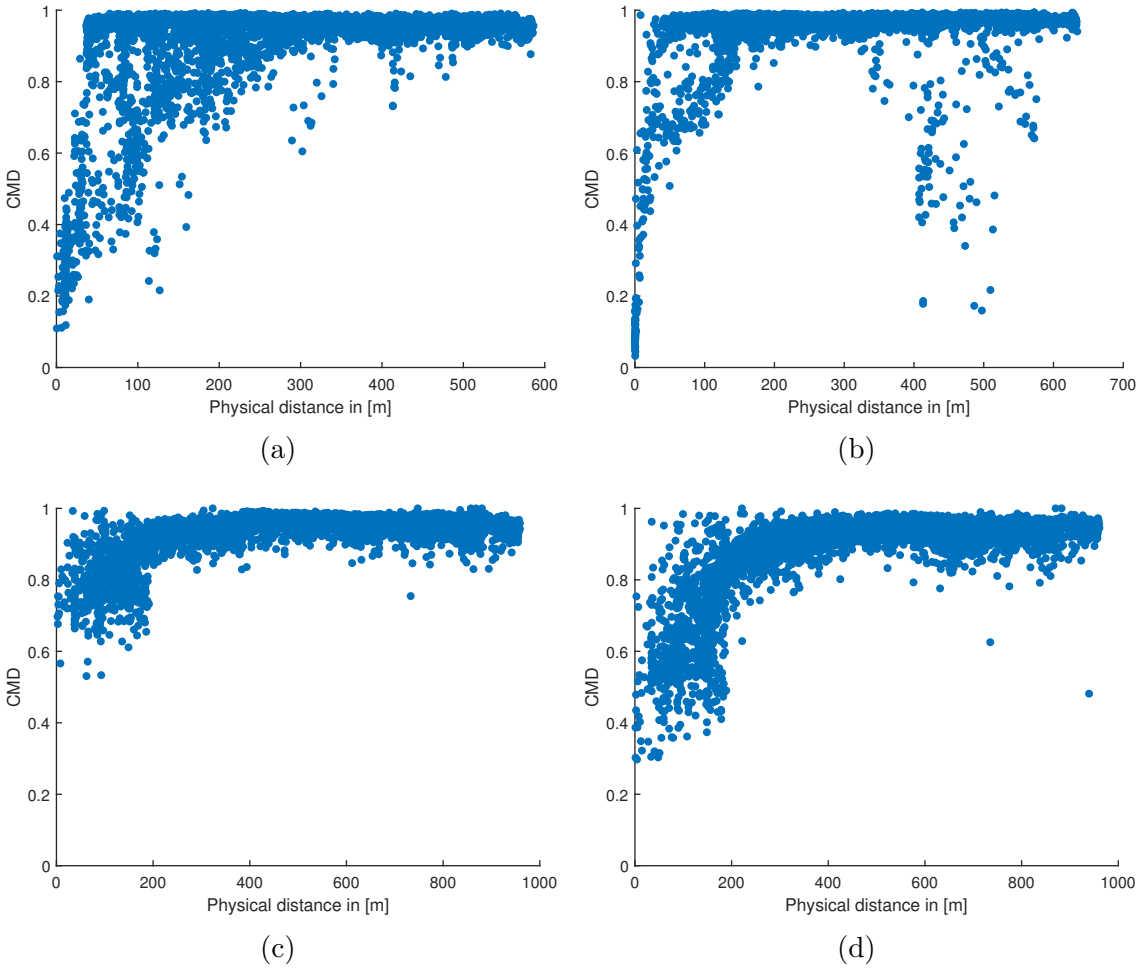
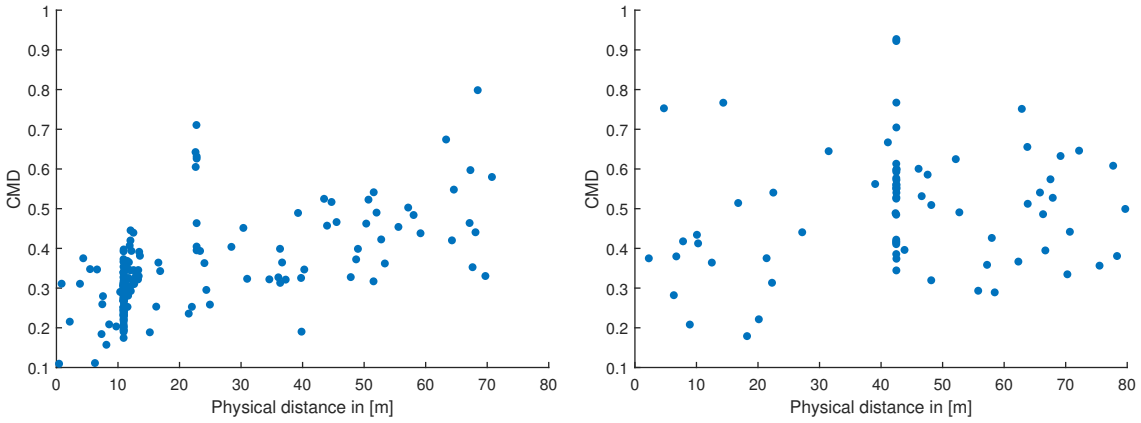


Figure 8: The relationship between the true distance and the CMD for 4 particular locations, compared to all other locations: (a); A LoS location. For low distances the dissimilarity depends on the distance, as expected. (b); A LoS location. In this case, there are distant UEs whose CSI appears similar to that of the depicted location. This phenomenon makes feature extraction and localization less accurate. (c); A NLoS location. There is little correlation between the true distance and the CMD, so features related to dissimilarities corresponding to this location will be of low quality. (d); A NLoS location. There is a higher correlation between true distance and CMD in this case, although not as strong as for the LoS locations. The difference between LoS and NLoS locations in terms of how accurately the dissimilarity can represent the true distance is clear. Hence, features based on dissimilarities will, in general, better represent the local geometry of transmitters in LoS areas, than that of transmitters in NLoS areas.

3.3.5 Inter-sample dissimilarities for LoS and NLoS locations

The effect of the existence of these two considerably distinct sets of locations, namely LoS and NLoS, with respect to the behavior of dissimilarities can be seen from an inter-sample perspective. That is, compare the samples at each location and calculate the



(a) (b)

Figure 9: The relationships between true distance and CMD for the neighboring points of two locations, in a way such that the rest of locations depicted are in the same general direction to it, close to a straight line. (a) LoS location. (b) NLoS location. The dependency of the CMD on the true distance is stronger for the LoS location. For the NLoS location, this relationship is hardly visible.

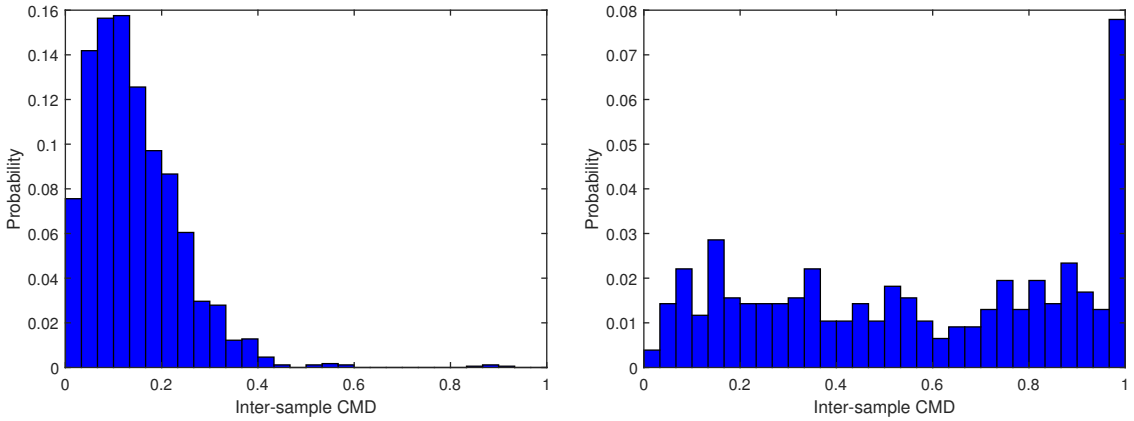
$\binom{S}{2}$—at most— pairwise dissimilarities between their associated covariance matrices.

Note that there might be less than $\binom{S}{2}$ defined inter-sample dissimilarities, since the dissimilarity between those samples with no information and any other sample is undefined. In other words, there exist dissimilarities $d_{\text{diss}}(\mathbf{R}_n, \mathbf{R}_m) = d_{\text{diss}}(\mathbf{R}_m, \mathbf{R}_n)$ that are undefined when one covariance matrix is $\mathbf{0}$.

Focusing now on those inter-sample dissimilarities which are defined, the effect of LoS and NLoS locations on the value and variance of these dissimilarities will be considerably different. Samples in NLoS locations can be considered to be more dissimilar than those in LoS locations. Therefore, features related to transmitter locations that have high inter-sample dissimilarities will be poorer representations of the said locations, when compared to locations in which the inter-sample dissimilarities are lower and less variant, as happens in LoS locations.

Having discussed the non-definiteness of dissimilarities when trivial covariance matrices are present, covariance matrices corresponding to samples with poor antenna reception will have several rows and columns equal to $\mathbf{0}$. Those samples will have, in general, high dissimilarities to the rest of the samples at the same location, as happened to dissimilarities between channels corresponding to locations which are in NLoS with the receiver, as discussed in 3.3.4.

Another consequence of this behavior is to do with utilizing an average of the samples for extracting features. In LoS locations, and since the observed channels for each of the samples will be highly self-similar, an average —either of the two discussed in 3.3.6— will be more representative of the true channel corresponding to those locations, i.e., the average between similar elements is also similar to the individual elements alone. For NLoS locations the opposite will occur; therefore the averaging in those locations will lead to poorer representations of the actual channel,



(a) (b)

Figure 10: Inter-sample CMD values (when defined) corresponding to two different sets of locations: (a) LoS locations. (b) NLoS locations. The difference between the two sets is clear. As expected, the values corresponding to LoS locations are relatively low. Note the high number of very high inter-sample dissimilarities for the NLoS set, due, in part, to the significant number of samples for which the covariance matrices are close to trivial.

since the mean has to be taken between samples corresponding to highly distinct channels.

Figure 10 illustrates these inter-sample dissimilarities for two sets of transmitters, namely sets located in LoS and NLoS locations. It is clear that covariance matrices of different samples at a particular location have higher similarities in LoS locations than in NLoS locations.

3.3.6 Averaging of samples

Due to the fact that there are 5 sets of measurements at every location, a logical way to combine these is by using the average of said measurements, so that the channel at one location can be represented by only one covariance matrix. The following two alternatives are considered:

1. The average of the channel values, so for each location the complex values are averaged, $\bar{\mathbf{h}}_{n,f} = \frac{1}{S}\sum_{s=1}^{S}\mathbf{h}_{n,:,f,s} \in \mathbb{C}^A$. After this averaging the covariance matrix is calculated as $\mathbf{R}_n = \mathbb{E}_f\left[\bar{\mathbf{h}}_{n,f}\bar{\mathbf{h}}_{n,f}^H\right]$.

2. The average of the covariance matrices corresponding to each of the five samples, $\mathbf{R}_n = \frac{1}{S}\sum_{s=1}^{S}\mathbf{R}_{n,s}$.

The differences between these two averaging approaches can be seen from the perspective of distances between the resulting covariance matrices (see section 3.3). Figure 11 illustrates these differences for the CMD. Table 4 contains the correlation coefficients of the dissimilarities obtained from covariance matrices using both averaging modes.

From this point of view, it can be concluded that there is a relatively small difference in using one or the other averaging approaches, as they give similar dissimilarities and, as a consequence, similar features. Thus, in this thesis only one averaging approach has been chosen, and the covariance matrix representing the channel from a transmitter location is the average of the covariance matrices corresponding to each of the samples.

In the following sections, $\mathbf{R}_n$ refers, in general, to the covariance matrix obtained from the CSI corresponding to location n, utilizing any of the averaging methods presented in this section—or any other.

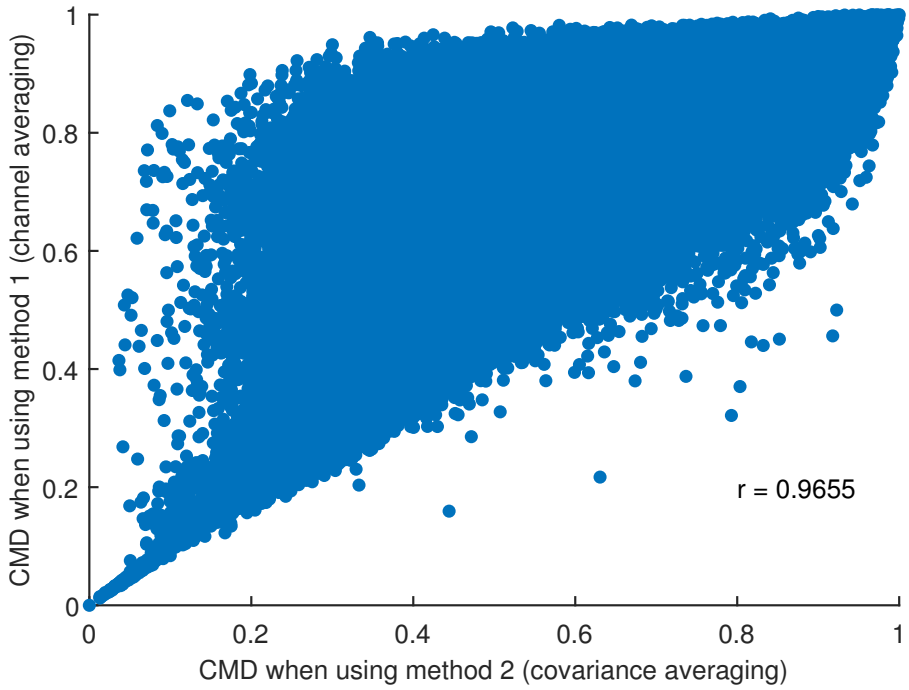


Figure 11: CMD when using both averaging methods, between all $\binom{N}{2}$ pairs of locations. r is the correlation coefficient.

Table 4: Correlation coefficient between dissimilarities (3.3) obtained using both averaging methods

Dissimilarity	Correlation coefficient
CMD	0.9655
Chordal	0.8083
X	0.9289
$\mathbb{C}\{\cdot\}$	0.8188
Signal subspace based	0.9449

3.4 Dimensionality reduction techniques

Once the channel parameters have been transformed to features with the methods discussed previously, each transmitter location is represented by a set of said features, which are in general of a high dimension. The objective of CC is to represent those features in a low-dimensional map which is able to capture the local geometry of the true transmitter locations. CC can be defined as a function C—not necessarily linear—that maps the channel features into the radio map,

$$C : \mathbb{C}^p \to \mathbb{R}^{p'}, \quad \text{where } p' < p. \tag{37}$$

For the task of obtaining a channel chart, several manifold learning and DR techniques have been used. The interest is in applying those algorithms that have the needed properties, such as neighborhood preservation. It is important to remark that the construction of a channel chart is performed in an unsupervised manner. That is, the DR algorithm finds representation of the spatial geometry of the true UE locations, but without any knowledge on where those locations are. All things considered, if the obtained channel features are able to represent the local geometry, the task of the dimensionality reduction is to find the intrinsic dimensionality of those features, which can be thought to be that of the spatial geometry.

3.4.1 Overview of dimensionality reduction algorithms

Several DR techniques have been used for the purpose of obtaining charts. This section contains a brief description of those techniques [34, 35].

For the description of the algorithms, the notation on Table 5 is used.

Table 5: Notation used in DR algorithm descriptions

$G = (\mathcal{V}, \mathcal{E})$	A graph G defined by set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E}$ connecting them.
$\mathbf{W}$	The weight matrix of a graph, in which $[\mathbf{W}]_{ij}$ is the weight of the edge connecting vertices i and j.
$\mathbf{D}$	The degree matrix of a graph, $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$, where $d_i = \sum_{j=1}^{N} [\mathbf{W}]_{ij}$.
$\mathbf{L}$	The unnormalized Laplacian matrix of a graph, given by $\mathbf{L} = \mathbf{D} - \mathbf{W}$.
$\mathcal{L}$	The normalized Laplacian matrix, given by $\mathcal{L} = \mathbf{D}^{-\frac{1}{2}}\mathbf{L}\mathbf{D}^{-\frac{1}{2}}$.
$\mathbf{1}$	A vector of ones of suitable size.
$\bar{\mathbf{x}}$	The mean of vector $\mathbf{x} \in \mathbb{R}^m$, i.e., $\bar{\mathbf{x}} = \frac{\mathbf{1}^T\mathbf{x}}{m}$.

3.4.2 Multidimensional Scaling

The goal of Multidimensional Scaling (MDS) [36, 37] is to find a low-dimensional spatial configuration of a set of data points with the only information of their general

Algorithm 1: Classical MDS

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$.
Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$.
1 Construct the matrix of squared proximities, $\mathbf{P}$, with $[\mathbf{P}]_{ij} = d^2_{\text{diss}}(\mathbf{x}_i, \mathbf{x}_j)$.
2 Compute the double-centering matrix $\mathbf{B} = -\frac{1}{2}\mathbf{JPJ}$, where $\mathbf{J} = \mathbf{I}_N - \frac{1}{N}\mathbf{1}\mathbf{1}^T$.
3 Find the eigendecompostion of $\mathbf{B}$ and order the eigenvalues in decreasing order, $\lambda_1 \geq \ldots \geq \lambda_N$. Label the corresponding eigenvectors accordingly, $\mathbf{v}_1, \ldots, \mathbf{v}_N$.
4 Construct $V_{p'} = [\mathbf{v}_1, \ldots, \mathbf{v}_{p'}]$ and $\mathbf{\Lambda}_{p'} = \text{diag}\,(\lambda_1, \ldots, \lambda_{p'})$.
5 The mapping onto the p'-dimensional embedding is $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T = \mathbf{V}_{p'}\mathbf{\Lambda}^{1/2}_{p'}$.

similarity—or dissimilarity, in a way such that the distances between data points are preserved as much as possible.

There exist multiple variants of the MDS algorithm. The variant mentioned and used in this thesis is Classical MDS, which is a linear DR technique. The steps of Classical MDS are detailed in Algorithm 1.

3.4.3 Laplacian Eigenmaps

Laplacian Eigenmaps (LE) [38] is a graph spectral technique which relies on the assumption that there is a knowledge of the structure of the underlying manifold in which the data is thought to reside.

LE has a neighborhood preserving characteristic which makes it relatively insensitive to the presence of outliers and noise. A consequence of this is the fact that the algorithm gives emphasis to natural clusters in the data. The steps of LE are shown in Algorithm 2.

3.4.4 Isomap

Isomap [39] was originally defined in order to solve the limitations of linear techniques, such as Classical MDS and PCA (Principal Component Analysis), in that they are unable to learn nonlinear manifolds. Isomap seeks to preserve the geometry of the data by approximating the geodesic distance between a pair of points in a way such that the distances of neighboring points are trusted to already be good approximations of the geodesic distance separating them, but for faraway points, geodesic distance is approximated by adding additional hops between neighboring points. These approximations utilize shortest path computations in a graph defined with the input data. The steps of Isomap are detailed in Algorithm 3.

3.4.5 Locally Linear Embedding

Locally Linear Embedding (LLE) [40, 41] is a DR algorithm in which the need to estimate pairwise distances —such as in Isomap— is eliminated. LLE attempts to find an underlying non-linear structure from locally linear fits. This way, a data

Algorithm 2: LE

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$.
Parameters: Number of nearest neighbors = K, heat kernel = T.
Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$.

1 Construct a weighted graph $G = (\mathcal{V}, \mathcal{E})$, where for two points $(\mathbf{x}_i, \mathbf{x}_j)$; $(i,j) \in \mathcal{E}$ if $\mathbf{x}_i$ is among the K nearest neighbors of $\mathbf{x}_j$, or vice-versa.
2 Construct a weight matrix $\mathbf{W}$, where $[\mathbf{W}]_{ij} = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{T}}$ if $(i,j) \in \mathcal{E}$, and $[\mathbf{W}]_{ij} = 0$ if $(i,j) \notin \mathcal{E}$. An unweighted graph is also possible, i.e., $[\mathbf{W}]_{ij} = 1$ if $(i,j) \in \mathcal{E}$, and $[\mathbf{W}]_{ij} = 0$ if $(i,j) \notin \mathcal{E}$.
3 Compute the diagonal weight (degree) matrix $\mathbf{D}$, and the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$.
4 Compute eigenvalues and eigenvectors of the generalized eigenvector problem $\mathbf{L}\mathbf{y} = \lambda \mathbf{D}\mathbf{y}$.
5 Order the eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{N-1}$, and label the eigenvectors $\mathbf{y}_0, \ldots, \mathbf{y}_{N-1}$ accordingly. Let $\mathbf{y}_s(n)$ be the n-th component of $\mathbf{y}_s$.
6 The image of point $\mathbf{x}_n$ onto the p'-dimensional embedding is given by $[\mathbf{y}_1(n), \ldots, \mathbf{y}_{p'}(n)]^T$. Note that $\mathbf{y}_0$ is discarded.

Algorithm 3: Isomap

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$.
Parameters: Number of nearest neighbors = K.
Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$.

1 Define a graph $G = (\mathcal{V}, \mathcal{E})$ such that $(i,j) \in \mathcal{E}$ if i is one of the K nearest neighbors of j according to a distance/dissimilarity d_{diss}.
2 Set the weight $w_{ij} = d_{\text{diss}}(\mathbf{x}_i, \mathbf{x}_j)$ if $(i,j) \in \mathcal{E}$, and 0 otherwise.
3 Compute the shortest paths between all pairs of points in G, and construct $\mathbf{D}_G$, containing the pairwise geodesic distances.
4 Construct matrices $\mathbf{S}_G$, where $[\mathbf{S}_G]_{ij} = [\mathbf{D}_G]_{ij}^2$, and the centering matrix $\mathbf{H} = \mathbf{I}_N - \frac{1}{N}\mathbf{1}\mathbf{1}^T$.
5 Find the eigendecomposition of $\frac{-\mathbf{H}\mathbf{S}_G\mathbf{H}}{2}$, order the eigenvalues in decreasing order $\lambda_1 \geq \ldots \geq \lambda_N$, and label the eigenvectors $\mathbf{y}$ accordingly. Let $\mathbf{y}_s(n)$ be the n-th component of $\mathbf{y}_s$.
6 The image of point $\mathbf{x}_n$ onto the p'-dimensional embedding is given by $\left[\sqrt{\lambda_1}\mathbf{y}_1(n), \ldots, \sqrt{\lambda_{p'}}\mathbf{y}_{p'}(n)\right]^T$.

point and its closest neighbors are considered to lie on a part of the manifold that is linear —or closely linear. The steps of LLE are described in Algorithm 4.

3.4.6 Local Tangent Space Alignment

Local Tangent Space Alignment (LTSA) [42] is defined as an improvement upon the capabilities of LLE. The idea behind LTSA is to assume a local linearity of

Algorithm 4: LLE

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$.
Parameters: Number of nearest neighbors $= K$.
Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$.

1 For each point $\mathbf{x}_i$, let $\mathcal{J}_i$ be the set containing its K-nearest neighbors.
2 Find the weight matrix $\mathbf{W}$ that minimizes the cost function given by $E(\mathbf{W}) = \sum_{i=1}^{N} \left\| \mathbf{x}_i - \sum_{j \in \mathcal{J}_i} w_{ij} \mathbf{x}_j \right\|^2, [\mathbf{W}]_{ij} = w_{ij}, \sum_j w_{ij} = 1, w_{ij} \geq 0$, by solving a contained least squares problem.
3 Find the output vectors $\mathbf{y}_1, \ldots, \mathbf{y}_N$, of dimension p', that minimize the embedding cost function given by $\Phi(\mathbf{Y}) = \sum_{i=1}^{n} \left\| \mathbf{y}_i - \sum_{j \in \mathcal{J}_i} w_{ij} \mathbf{y}_j \right\|^2$, subject to $\mathbf{Y}\mathbf{Y}^T = \mathbf{I}_{p'}$, where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$, by solving a generalized eigenvalue problem.

the manifold, which defines a linear mapping from a data point to its local tangent space, and a linear mapping from the corresponding low-dimensional data point to said tangent space. In aligning these local tangent spaces, the global space of the manifold can be constructed. This way, the process of constructing a non-linear manifold is combined with the process of reconstruction of the said manifold. The steps of LTSA are described in Algorithm 5.

Algorithm 5: LTSA

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$.
Parameters: Number of nearest neighbors $= K$.
Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$.

1 **for** $n = 1, \ldots, N$ **do**
2 Find the K nearest neighbors of $\mathbf{x}_n$, $\mathbf{x}_{n_k}$, for $k = 1, \ldots, K$, and let $\mathbf{X}_n = [\mathbf{x}_{n_1}, \ldots, \mathbf{x}_{n_K}]$.
3 Compute the eigenvectors of $\left(\mathbf{X}_n - \bar{\mathbf{x}}_n \mathbf{1}^T\right)^T \left(\mathbf{X}_n - \bar{\mathbf{x}}_n \mathbf{1}^T\right)$, and select the eigenvectors $\{\mathbf{v}_l\}_{l=1}^{p'}$ corresponding to the p' largest eigenvalues.
4 Construct the matrix $\mathbf{G}_n = \left[\frac{\mathbf{1}}{\sqrt{K}}, \mathbf{v}_1, \ldots, \mathbf{v}_{p'}\right] \in \mathbb{R}^{K \times (p'+1)}$.
5 Compute $\mathbf{B}_n = \mathbf{I}_K - \mathbf{G}_n \mathbf{G}_n^T \in \mathbb{R}^{K \times K}$.
6 **end**
7 Construct the block diagonal matrix $\mathbf{B} = \text{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_N) \in \mathbb{R}^{NK \times NK}$.
8 Sort the eigenvalues of $\mathbf{B}$ in ascending order, $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{NK-1}$, and label the corresponding eigenvalues $\mathbf{y}$ accordingly. Let $\mathbf{y}_s(n)$ be the i-th component of $\mathbf{y}_s$.
9 The image of point $\mathbf{x}_n$ onto the p'-dimensional embedding is given by $[\mathbf{y}_1(n), \ldots, \mathbf{y}_{p'}(n)]^T$. Note that $\mathbf{y}_0$ is discarded.

3.4.7 t-SNE

t-SNE (short for t-Distributed Stochastic Neighbor Embedding) is a variation of SNE [43, 44]. These DR techniques are based on converting the distances between data points into similarities given by conditional probabilities related to local neighborhoods.

t-SNE aims to solve some of the problems of SNE by using probabilities based on Student t-distributions and a cost function with simpler gradients. The steps of t-SNE are shown in Algorithm 6.

Algorithm 6: t-SNE

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$.
Parameters: Perplexity $= Perp$, learning rate η, momentum $\alpha(t)$, number of iterations T.
Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$.

Compute pairwise similarities $p_{m|n} = \frac{\exp\left(-\|\mathbf{x}_n - \mathbf{x}_m\|^2\right)/2\sigma_n^2}{\sum_{k \neq n} \exp\left(-\|\mathbf{x}_n - \mathbf{x}_k\|^2\right)/2\sigma_n^2}$, with a value for σ_n that satisfies $Perp = 2^{-\sum_m p_{m|n} \log_2 p_{m|n}}$.

Set $p_{nm} = \frac{p_{n|m} + p_{m|n}}{2N}$.

Sample an initial solution for $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{p'}$ from $N(0, 10^{-4}\mathbf{I})$ or use a predefined initial solution from other DR techniques. Let $\mathcal{Y}^{(0)} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ be the initial solution.

for $t = 1, \ldots, T$ **do**

  Compute the low-dimensional affinity $q_{nm} = \frac{\left(1 + \|\mathbf{y}_n - \mathbf{y}_m\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2\right)^{-1}}$.

  Find the cost gradient $\frac{\partial C}{\partial \mathcal{Y}}$, with the cost function C defined in terms of the Kullback-Leibler divergence, as $C = \sum_{n=1}^{N} \mathrm{KL}(P_n \| Q_n) = \sum_{n=1}^{N} \sum_{m=1}^{N} p_{nm} \log \frac{p_{nm}}{q_{nm}}$.

  Update $\mathcal{Y}$ with gradient descent: $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta \frac{\partial C}{\partial \mathcal{Y}} + \alpha(t) \left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)}\right)$.

end

The result is found by $\mathcal{Y}^{(T)}$

3.4.8 Effect of using non-labeled data for CC

Since the dataset contains non-labeled data points, it can be useful to study the effect of utilizing it for channel charting. This can be done with the dimensionality reduction algorithms of 3.4, since they are unsupervised and, as such, do not require any sort of side information. The addition of data to be used for channel charting can be an advantage since the feature geometry can become a more accurate representation of the spatial geometry of the transmitter locations.

In regard to the scores to characterize features and charts described in 3.2, it is important to emphasize that said scores can only be computed from the perspective

of the known locations, as they are a comparison between the spatial geometry, i.e. the true locations, and the feature or chart geometry.

3.5 Graph signal processing

There exist algorithms to interpolate data defined on graphs using signal processing techniques. These algorithms fall into the domain of semi-supervised learning. Contrary to the algorithms described in 3.4, in this case the unknown location of missing points is found by using true location information.

It is important to describe how graph signal processing is seen from the perspective of dimensionality reduction and localization. The location for some points is known, and this information is fetched into the algorithm, which combines it with the feature information used in the algorithms in section 3.4. In this sense, it can be seen as a semi-supervised dimensionality reduction algorithm. Since some locations are known, it is expected that the mapping of points on the low-dimensional manifold will be a better representation than in the unsupervised techniques already described. Depending on the amount of data points for which the location is known, the algorithm will yield better or worse results.

The interpolation algorithm used in this thesis is described in Algorithm 7, based on [45].

3.6 Localization

This section contains a description of the localization methods that have been utilized in this thesis.

After having found features and charts which, in a way, contain information about the true location of the transmitters, this information can be used to locate new UEs, both from the feature perspective and by using the constructed channel charts.

There are, therefore, two different ways to approach localization. In the first one, the calculated features or dissimilarities are used to estimate new locations. Secondly, a channel chart constructed from those features can be used for localization.

Since the dataset contains true locations for a set of transmitters (those labeled data points), they can be used to evaluate the accuracy of location estimations, by comparing both true and estimated locations.

It is worthwhile to emphasize the nature of the particular localization procedure that is done in this thesis. As a typical machine learning problem, the dataset can be divided into training and test sets. For this particular case, the training set contains data points whose location is known, whereas the test set contains data points for which the location is estimated—with the localization methods described in this section—, and then compared to their true location for performance evaluation.

Table 6 includes the notation utilized in this section to describe the different localization approaches.

Algorithm 7: Graph Interpolation Method

Data: N data points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^p$. True locations (of dimension p') of a set of data points, $\mathbf{Y}_{\mathcal{S}} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_{|\mathcal{S}|}\right]^T$.

Parameters: Heat kernel $= T$, θ

Result: N data points $\mathbf{y}_1, \ldots, \mathbf{y}_N \in \mathbb{R}^{d'}$.

// Let $\mathcal{S}$ be the set of labeled data and $\mathcal{S}^C$ the set of unlabeled data. A matrix $\mathbf{A}_{\mathcal{S}}$ denotes a submatrix of $\mathbf{A}$ indexed with the indices of the rows and/or columns of $\mathcal{S}$, and analogously for $\mathcal{S}^C$.

1 Construct a weighted graph $G = (\mathcal{V}, \mathcal{E})$, where for two points $(\mathbf{x}_i, \mathbf{x}_j); (i,j) \in \mathcal{E}$ if $\mathbf{x}_i$ is among the K-nearest neighbors of $\mathbf{x}_j$, or vice-versa, according to a distance or dissimilarity d_{diss}.

2 Construct a weight matrix $\mathbf{W}$, with $[\mathbf{W}]_{ij} = d_{\text{diss}}(\mathbf{x}_i, \mathbf{x}_j)$ if $(i,j) \in \mathcal{E}$, and 0 otherwise.

3 Compute the diagonal weight (degree) matrix $\mathbf{D}$, and the unnormalized Laplacian $\mathbf{L}$.

4 Let $\omega_{\mathcal{S}}$ be the square-root of smallest eigenvalue of the submatrix corresponding to the unlabeled data of the squared normalized Laplacian, $\mathcal{L}^2_{\mathcal{S}^C}$.

5 Update the weights between labeled data points, $\left[\hat{\mathbf{W}}\right]_{ij} = [\mathbf{W}]_{ij} \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\theta^2}\right) \forall \{i,j \in \mathcal{S}\}$.

6 Find the normalized Laplacian matrix $\mathcal{L}$ and the degree matrix $\mathbf{D}$ using the updated weights from $\hat{\mathbf{W}}$.

7 Find the eigendecomposition of $\mathcal{L}$ and order its eigenvalues in ascending order, $\lambda_1 \leq \ldots \leq \lambda_N$. Label the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_N$ accordingly.

8 Let $b = \arg\max_b \left(\lambda_b < \omega_{\mathcal{S}} \mid b = 1, \ldots, N\right)$.

9 Construct $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_b]$. Make $\mathbf{V}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}| \times b}$ and $\mathbf{V}_{\mathcal{S}^C} \in \mathbb{R}^{|\mathcal{S}^C| \times b}$ by taking the corresponding columns of $\mathbf{V}$.

10 Compute $\mathbf{G}_{\mathcal{S}} = \mathbf{D}_{\mathcal{S}}{}^{\frac{1}{2}} \mathbf{Y}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}| \times p'}$ and $\mathbf{G}_{\mathcal{S}^C} = \mathbf{V}_{\mathcal{S}^C} \left(\mathbf{V}_{\mathcal{S}}^T \mathbf{V}_{\mathcal{S}}\right)^{-1} \mathbf{D}_{\mathcal{S}}{}^{\frac{1}{2}} \mathbf{Y}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}^C| \times p'}$.

11 The interpolated points are given by $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T = \mathbf{D}^{-\frac{1}{2}} \mathbf{G}$, where $\mathbf{Y}_{\mathcal{S}^C}$ contains the estimates of the true locations of the unlabeled data points.

Table 6: Notation related to localization

$\mathbf{x}_n \in \mathbb{C}^p$	Vector of features of a data point n.
$\mathbf{y}_n \in \mathbb{R}^{p'}$	True location of a data point n.
$\hat{\mathbf{y}}_n \in \mathbb{R}^{p'}$	Location estimation of a data point n.
$\mathcal{T} = \left\{t_1, \ldots, t_{\lvert\mathcal{T}\rvert}\right\}$	Indices of the data points corresponding to the training set, or labeled data.
$\mathcal{T}^C = \left\{t_1^C, \ldots, t_{\lvert\mathcal{T}^C\rvert}^C\right\}$	Indices of the data points corresponding to the test set, or unlabeled data.
$\mathcal{X}_\mathcal{T} = \left\{\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_{\lvert\mathcal{T}\rvert}}\right\}$	Features of points in the training set.
$\mathcal{X}_{\mathcal{T}^C}$	Features of points in the test set.
$\mathcal{Y}_\mathcal{T} = \left\{\mathbf{y}_{t_1}, \ldots, \mathbf{y}_{t_{\lvert\mathcal{T}\rvert}}\right\}$	Locations of points in the training set.
$\mathcal{Y}_{\mathcal{T}^C}$	Locations of points in the test set.
$\mathcal{G}_{K_n} = \{\mathbf{g}_{n_1}, \ldots, \mathbf{g}_{n_K}\} \subset \mathcal{X}_\mathcal{T}$	Feature vectors of the K nearest points of $\mathbf{x}_n$ in the labeled set, according to a dissimilarity d_{diss} or related to it, indexed from nearest to furthest.
$\mathcal{S}_{K_n} = \{\mathbf{s}_{n_1}, \ldots, \mathbf{s}_{n_K}\} \subset \mathcal{Y}_\mathcal{T}$	True locations of the K nearest points of $\mathbf{x}_n$ in the labeled set, according to a dissimilarity d_{diss} or related to it, indexed from nearest to furthest.

The dissimilarity d_{diss} can also be defined with respect to covariance matrices that represent the data points $\mathbf{x}$.

3.6.1 KNN methods

Localization by K-Nearest Neighbors (KNN) [19, 33] takes into account the fact that the feature geometry is a representation of the true location of the transmitters. Furthermore, if local neighborhood preservation is assumed, those features can be utilized to perform KNN localization.

All KNN methods used in this thesis share the same principle. For estimating the position of a UE for which CSI has been collected, dissimilarities to the channels corresponding to all known locations are found. The estimated location is a combination of the K-nearest neighbors, in terms of a chosen dissimilarity, with K being a tuneable parameter. The difference in the various KNN methods lies in how these dissimilarities are utilized to find the estimation. Three of these combinations are considered in this thesis.

- **KNN with pairwise distances**

  This approach is the simplest form of KNN localization.

  For a query data point $\mathbf{x}_q \in \mathcal{X}_{\mathcal{T}^C}$, the dissimilarities to all data points in the training set, $\mathcal{X}_\mathcal{T}$, are found, $d_{\text{diss}}(\mathbf{x}_q, \mathbf{x}_i), \forall i \in \mathcal{T}$. These dissimilarities are

ordered in an increasing order, and the data points which constitute the K lowest dissimilarities are the elements of $\mathcal{G}_{K_q}$, with corresponding locations $\mathcal{S}_{K_q}$.

The estimated location of $\mathbf{x}_q$, $\hat{\mathbf{y}}_q$ is given by the barycenter of the locations corresponding to the K-nearest neighbors, $\hat{\mathbf{y}}_n = \frac{1}{K}\sum_{i=1}^{K}\mathbf{s}_{q_i}$. For the particular case of $K = 1$, the location estimation is equal to the location of the nearest neighbor of $\mathbf{x}_q$, i.e., $\hat{\mathbf{y}}_q = \mathbf{s}_{q_1}$.

- **KNN with dissimilarity vectors**

  This approach differs from the previous one in that not only the $|\mathcal{T}|$ dissimilarities between the query point $\mathbf{x}_q \in \mathcal{X}_{\mathcal{T}^C}$ and all the points in $\mathcal{X}_{\mathcal{T}}$ are found, but also all $\binom{|\mathcal{T}|}{2}$ pairwise dissimilarities among the points in the training set.

  For every data point a vector of dissimilarities to all points in the training set can be defined, $\mathbf{v}_i = \left[d_{\text{diss}}(\mathbf{x}_{t_1}, \mathbf{x}_i), \ldots, d_{\text{diss}}(\mathbf{x}_{t_{|\mathbf{T}|}}, \mathbf{x}_i)\right]^T$.

  For a query point $\mathbf{x}_q \in \mathcal{X}_{\mathcal{T}^C}$ its corresponding vector $\mathbf{v}_q$ is found, and the vector distances between this vector and the ones corresponding to the points in the labeled set, $\mathcal{D}_q = \{\|\mathbf{v}_q - \mathbf{v}_i\| \mid i \in \mathcal{T}\}$. The location estimation is done with the locations corresponding to the elements of $\mathcal{D}_q$ with the K lowest values, $\mathcal{S}_{K_q}$. The estimation $\hat{\mathbf{y}}_q$ is performed in the same way as explained in the previous KNN method, i.e., $\hat{\mathbf{y}}_q = \frac{1}{K}\sum_{i=1}^{K} s_{q_i}$.

- **KNN with inverse dissimilarity vectors**

  A situation that can arise with the use of the previously defined dissimilarity vectors, given how they are constructed, is that the value of vector distances mainly depends on high differences in dissimilarities, which might correspond to points that are far away and, therefore, should influence less in the finding of the nearest neighbors, compared to those who are close. The relationship between dissimilarity and distance has been previously discussed in section 3.3.4, and the fact that high dissimilarities are unable to represent distances between points has been observed. Therefore, high dissimilarities can, in general, be less trusted than those who are relatively low, when it comes to representing true distances and, in turn, nearest neighbors.

  In order to counteract the effect that high dissimilarities have a great impact in the way neighbors are determined —and low dissimilarities have a lesser impact— it can be useful to define a new distance vector in which the low —and more trustworthy— dissimilarities are given a greater weight. One natural way to achieve this is to define a dissimilarity vector $\mathbf{v}_i$ between point $\mathbf{x}_i$ and the points in the training set, $\mathbf{v}_i = \left[d^{-\alpha}_{\text{diss}}(\mathbf{x}_{t_1}, \mathbf{x}_i), \ldots, d^{-\alpha}_{\text{diss}}(\mathbf{x}_{\mathbf{t}_{|\mathcal{T}|}}, \mathbf{x}_i)\right]^T$, for a chosen value of $\alpha \in \mathbb{R}^*_+$.

  For the entries of $\mathbf{v}_i$ where the corresponding dissimilarity d_{diss} is 0, i.e., when it is calculated between a data point and itself, that entry is manually modified in a way such that it does not contribute to the distance $\|\mathbf{v}_q - \mathbf{v}_i\|$, to avoid

indeterminations. In other words, entries of $\mathbf{v}_i$ which are zero are replaced by the entries of $\mathbf{v}_q$ at the same positions.

From this point the localization procedure with inverse dissimilarity vectors follows analogously that of the previous approach.

It can be useful to define the location estimations in terms of a weighted average, instead of using the barycenter of the locations of the K-nearest neighbors. This way, it is possible to define a weighted estimation in which the K neighbors are given different weights, which are inversely proportional to their related dissimilarities. This way, the location estimation $\hat{\mathbf{y}}_q$ for a point $\mathbf{x}_q \in \mathcal{X}^{\mathcal{T}^C}$ can be found by defining the weights $\mathbf{w}_{K_q} = [w_{q_1}, \ldots, wq_K]^T = [d^{-\gamma}(\mathbf{x}, \mathbf{g}_{q_1}), \ldots, d^{-\gamma}(\mathbf{x}, \mathbf{g}_{q_K})]^T$, for $\gamma \geq 0$.

$$\hat{\mathbf{y}}_q = \frac{1}{K \sum_{i=1}^{K} w_{q_i}} \sum_{j=1}^{K} w_{q_j} \mathbf{s}_{q_j}. \tag{38}$$

Note that $\gamma = 0$ implies no weighting, and the estimation is the barycenter of the K locations, as explained previously.

To study the performance of localization, for all of the three approaches previously described, the mean error in [m] is used. For estimations $\hat{\mathbf{Y}}_{\mathcal{T}^C} = \left[\hat{\mathbf{y}}_{t_1^C}, \ldots, \hat{\mathbf{y}}_{t^C_{|\mathcal{T}^C|}}\right]$, the average localization error will simply be

$$e = \frac{1}{|\mathcal{T}^C|} \sum_{i \in \mathcal{T}^C} \|\hat{\mathbf{y}}_i - \mathbf{y}_i\| . \tag{39}$$

Another useful manner to analyze the localization error is to look at its empirical cumulative density function (CDF). This function provides an insight into the distribution or errors, and it can be of use to analyze it alongside the average error.

It is worthwhile to further analyze the localization methods and comment on properties that can help understand the differences among them.

An important observation to make is that localization results depend on the density of points in the training set, i.e., the distances between their true locations. A dense set of points will, in general, yield more accurate results than a set in which the pairwise distances are relatively high, given that the estimates are found by a combination of locations in the training set.

An essential difference between KNN with pairwise distances and the dissimilarity vector approaches is the effect of updates of the training set in determining the neighbors of a query point. In the case of dissimilarity vectors, if the training set is updated with the inclusion of a new element, the dissimilarity vectors need to also be updated, and all vector distances in $\mathcal{D}$ need to be computed again. Therefore, the inclusion of a new element in the training set can change the locations corresponding to its K lowest values in an uncertain manner. This situation does not occur when localization is done with using pairwise dissimilarities. In this case, the inclusion of a new element $\mathbf{x}_r$ in $\mathcal{X}_{\mathcal{T}}$ will give an updated set $\mathcal{S}'_{K_q}$, for which the elements corresponding to the K-nearest neighbors will be $\mathcal{S}'_{K_q} = \mathcal{S}_{K_q}$ if $d_{\text{diss}}(\mathbf{x}_q, \mathbf{x}_{q_K}) > d_{\text{diss}}(\mathbf{x}_q, \mathbf{x}_r)$ or $\mathcal{S}'_{K_q} = \left\{\mathbf{s}_{q_1}, \ldots, \mathbf{s}_r, \ldots, \mathbf{s}_{q_{K-1}}\right\}$ otherwise. Thus, the

K-nearest neighbors after the training set update will either be the $K-1$-nearest neighbors from before the update plus the new element, or there will be no change.

3.6.2 Extreme Learning Machine

An Extreme Learning Machine (ELM) is a single-hidden layer feedforward neural network (SLFN) that can be used for classification, regression and other deep learning applications [46]. The particularity of an ELM is that the parameters of its hidden layers need not be learned, and the learnable parameters are found analytically to speed up their learning, contrary to regular neural networks which use gradient-based learning and weight update.

For the particular purpose of localization, ELMs can be utilized as a supervised learning approach. In the case of this thesis, a simple ELM consisting of one hidden layer has been used [33].

An ELM can be thought of as a non-linear function, given by

$$\text{ELM} : \mathbf{x}_n \in \mathbb{R}^p \rightarrow \hat{\mathbf{y}}_n = \boldsymbol{\beta}\varphi(\mathbf{W}^T\mathbf{x}_n) \in \mathbb{R}^{p'}, \tag{40}$$

where $\mathbf{x}_n$ are the input features, $\varphi(\cdot)$ is a non-linear activation function, $\mathbf{W} \in \mathbb{R}^{p\times\mu}$ is the weight matrix (for μ hidden neurons), and $\boldsymbol{\beta} \in \mathbb{R}^{p'\times\mu}$ are the output weights. In this thesis, the considered activation function is the Rectified Linear Unit (ReLU), i.e., $\varphi(x) = \max(0, x)\ \forall x \in \mathbb{R}$.

In the training phase, the entries of the matrix $\mathbf{W}$ are drawn from a Gaussian standard distribution, i.e., $[\mathbf{W}]_{ij} \sim N(0, 1)$, and remain constant. The only learned parameters are the output weights $[\boldsymbol{\beta}]_{ij}$.

Let $\mathcal{T}$ and $\mathcal{T}^C$ be the indices representing the training and test datasets, respectively. From the perspective of localization, $\mathbf{Y}_\mathcal{T} \in \mathbb{R}^{p'\times|\mathcal{T}|}$ are the true locations. Let $\boldsymbol{\Sigma}_\mathcal{T} = \varphi\left(\mathbf{W}^T\mathbf{X}_\mathcal{T}\right) \in \mathbb{R}^{\mu\times|\mathcal{T}|}$ denote the hidden neuron activations, where $\mathbf{X}_\mathcal{T} = \left[\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_{|\mathcal{T}|}}\right]$ contains the features from the training set.

There exist numerous ways of defining the feature vector $\mathbf{x}$. In this thesis, the vector corresponding to a data point contains pairwise dissimilarities to all the other points in the training set, for a given dissimilarity.

The output weight matrix $\boldsymbol{\beta}$ is found by solving a simple least squares linear regression problem, and is given by

$$\boldsymbol{\beta} = \mathbf{Y}_\mathcal{T}\left(\boldsymbol{\Sigma}_\mathcal{T}^T\boldsymbol{\Sigma}_\mathcal{T} - \gamma\mathbf{I}_{|\mathcal{T}|}\right)^{-1}\boldsymbol{\Sigma}_\mathcal{T}^T, \tag{41}$$

where γ is an L_2 regularization parameter to prevent the network from overfitting and performing poorly on the test set, or on other query points.

The location estimations $\hat{\mathbf{Y}}_{\mathcal{T}^C}$ for features corresponding to the test set are found using (40), i.e., $\hat{\mathbf{Y}}_{\mathcal{T}^C} = \left[\hat{\mathbf{y}}_{t_1^C}, \ldots, \hat{\mathbf{y}}_{t^C_{|\mathcal{T}^C|}}\right] = \boldsymbol{\beta}\boldsymbol{\Sigma}_{\mathcal{T}^C}$. The average localization error of the estimations $\hat{\mathbf{Y}}_{\mathcal{T}^C}$ is computed using (39).

3.6.3 Localization based on CC

For the localization methods introduced in the previous section, namely KNN and ELM, it is crucial to emphasize that these methods can be utilized in two different

ways. Firstly, and as already described, localization can be performed directly on the features. That is, with the use of dissimilarities in the ways described, and without using a channel chart. Secondly, localization based on channel charts—which are constructed from the features—can be also considered.

A logical motivation for using channel charts for localization is to understand that they constitute a representation of the spatial geometry, in which local neighborhood relationships are preserved. Thus, and since this property is the same that motivates dissimilarity based localization, one can use the same methods for performing localization using radio charts instead of dissimilarities like those introduced in section 3.3. The new dissimilarities can simply be thought of as the distances between the points in the chart, and the same principles from sections 3.6.1 and 3.6.2 can be applied.

4 Results

This chapter includes a description of the results obtained with the tools and methods contained in the previous chapter. It is divided in three parts; in the first part the results related to CC are presented, the second part includes a description of the results related to localization. The final part includes a brief summary of the results obtained with the graph signal processing algorithm presented in section 3.5.

4.1 CC results

In this section, a portrayal of the results concerning CC is included. The main method for evaluating these results are the quality scores introduced in section 3.2. As previously discussed, these scores can be used not only to evaluate the quality of a generated radio chart, but to evaluate the quality of the features that have been used to construct the mentioned chart. A logical way of analyzing the CC results is to first look at the scores of the features, and then compare them to the scores of the generated charts. The usefulness of charts—at least for the applications concerned in this thesis—can be analyzed from this point of view. That is, the interest lies in finding charts whose geometry is able to represent the spatial geometry of the true locations in a more accurate way than the geometry of the features. If a particular chart has significantly poorer scores than an analyzed set of features, it is of little use to exploit it for gaining knowledge of the original spatial geometry, or to use it for localization. However, it is important to remark that the usefulness of channel charts is also related to the fact that they are low-dimensional. Thus, finding features with excellent scores is not enough, and the focus must be in learning charts which have those high scores, higher than those of the features, if possible.

The features studied in this thesis come from the exploitation of the dissimilarities presented in section 3.3. That is, every data point $\mathbf{x}_n$ contains the dissimilarity values to all other points:

$$\mathbf{x}_n = \left[d_{\text{diss}}\left(\mathbf{R}_n, \mathbf{R}_1\right), \ldots, d_{\text{diss}}\left(\mathbf{R}_n, \mathbf{R}_N\right)\right]^T. \tag{42}$$

It is important to remark that utilizing features based on dissimilarities is not a requisite for constructing charts; other features—usually related to covariance matrices—can also be used.

Table 7 shows the TW, CT and KS scores (see section 3.2) corresponding to the dissimilarity features utilized in this thesis. For TW and CT, the scores are given for different values of the number of neighbors, i.e., $K = \{10, 25, 50, 100\}$. Note that these two scores, by definition, necessarily decrease—or remain constant—as K increases.

Once the quality of the features has been found, the next logical step is to, by using the DR techniques detailed in section 3.4, construct charts that are able to improve the scores of the features, presented on Table 7.

There is a significant number of combinations of features and DR techniques —some with tuneable parameters—for constructing charts. However, presenting the results of all of these combinations does not add significant value to the discussion

Table 7: Dissimilarity feature scores

d_{diss}	KS	K							
		10		25		50		100	
		TW	CT	TW	CT	TW	CT	TW	CT
CMD	0.480	0.944	0.934	0.939	0.933	0.935	0.932	0.930	0.929
Chordal	0.501	0.930	0.908	0.921	0.906	0.914	0.903	0.906	0.900
X	0.527	0.926	0.904	0.913	0.900	0.900	0.892	0.880	0.874
$\mathbb{C}\{\cdot\}$	0.452	0.942	0.898	0.936	0.897	0.930	0.895	0.924	0.887
S. sub.	0.490	0.945	0.932	0.939	0.932	0.934	0.930	0.929	0.926

and analysis of the results obtained. Therefore, only those results which are relevant and worth discussing are presented. It is patent that, for the particular dataset on which this thesis is based, there are certain of these combinations that provide better results than others. From Table 7, it is clear that not all dissimilarities have the same quality. In particular, the features obtained by means of the chordal and X distances are, in general terms, worse representations of the true locations of the transmitters, from the perspective of the quality scores in question. As a result, some of the results that are presented in this section only focus on the remaining 3 dissimilarities.

The generated charts can be not only evaluated by means of the quality scores; in addition they can be evaluated visually. The generated channel charts will be p'-dimensional, in general. By choosing a value of p' that allows for visualization, i.e. 2 or 3, the chart can be visualized. In this thesis a value of $p' = 2$ has been chosen. Moreover, having charts with dimensions as low as possible allows for a more efficient learning of the spatial geometry at the receiver.

The choice of p' can furthermore be justified from the point of view of the spatial geometry of the transmitter locations, since they lie on a plane, after having removed the height dimension, which had little variance compared to the x and y coordinates (see section 3.1.1). The visual interpretation of the channel charts can be made by comparing them to a plot representing the true locations in which the neighborhoods are emphasized. Figure 12 illustrates this idea. The true locations of the transmitters can be seen, and points which are in the neighborhood of one another have similar colors. A good channel chart will have an appearance similar to that of this figure, without taking into account aspects such as rotation, scaling, translation or reflection, since these transformations do not affect neighborhood preservation, and the quality scores in question remain unaffected by them (see section 3.2).

Table 8 contains the quality scores of charts obtained with the DR algorithms previously detailed. The features utilized for the construction of these charts are, again, the dissimilarity features obtained from equation (42), with the following exceptions: for LLE, instead of the dissimilarity features, the features for $\mathbb{C}\{\cdot\}$ is calculated by normalizing the real and imaginary parts of the covariance matrix (vectorized, as in eq. (24), without the $\sqrt{2}$ terms) by the normalization term defined in equation (15). For the signal subspace dissimilarity, the feature vectors are

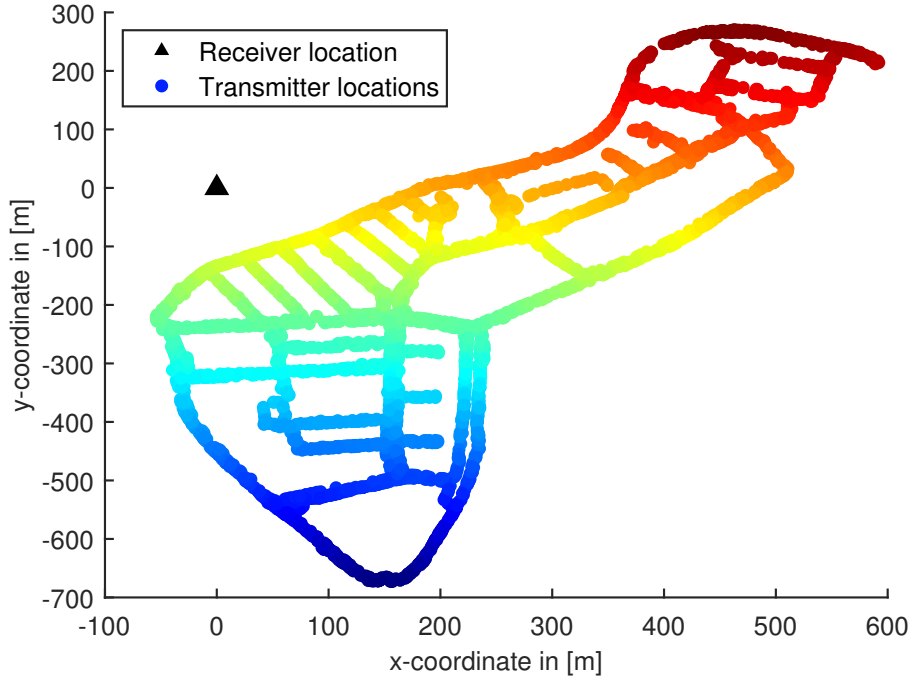


Figure 12: Receiver and transmitter locations. Nearby points have similar coloring.

Table 8: Scores for the charts generated with the DR algorithms detailed in section 3.4, for the specified dissimilarities, using the dissimilarity features defined in equation (42). TW and CT are computed with a value of $K = 50$. The parameters used for the DR algorithms are: LE: $T = 0.1, K = 248$. Isomap: $K = 8$, LTSA & LLE: $K = 248$, t-SNE: Initial solution = sampled from a random distribution, $Perp = 200$, $T = 1000$, $\eta = 500$, $\alpha(t) = 0.5$ for $t < 250$ and 0.8 for $t \geq 250$

DR method	CMD			$\mathbb{C}\{\cdot\}$			S. Sub.		
	KS	TW	CT	KS	TW	CT	KS	TW	CT
MDS	0.466	0.842	0.892	0.462	0.850	0.867	0.438	0.818	0.872
LE	0.466	0.902	0.930	0.391	0.890	0.911	0.491	0.896	0.921
Isomap	0.446	0.861	0.911	0.382	0.872	0.911	0.464	0.841	0.915
LLE*	0.574	0.732	0.770	0.673	0.615	0.764	0.546	0.746	0.822
LTSA	0.550	0.853	0.889	0.622	0.835	0.844	0.548	0.857	0.901
t-SNE	0.395	0.926	0.944	0.311	0.931	0.940	0.366	0.927	0.938

$\mathbf{x}_n = [f_{n1}, \ldots, f_{nN}]$, following the similarity defined in eq. (21). It can be observed that the charts generated directly with the DR algorithms in question are worse representations of the spatial geometry than the dissimilarity features, with the exception of t-SNE, which produces charts of higher quality. This fact requires other combinations—and potentially other DR techniques— to be explored, in order to find charts with better quality scores.

A combination that has led to charts whose scores have been better than those of the dissimilarities is using the bi-dimensional result obtained from the Isomap

algorithm, when using a small value for its parameter K_{Iso}, as a pre-defined initial solution to t-SNE, instead of sampling it from a random distribution (see Algorithm 6). The resulting charts are shown in Figure 13, and its corresponding quality scores (for values of $K = \{10, 25, 50, 100\}$) are shown on Table 9. These charts constitute, by terms of their quality scores, the most accurate representations of the spatial geometry that have been achieved throughout the development of this thesis. It is worth noticing that these charts have, in general, better scores than those of the dissimilarity related features found on Table 7. Therefore, it can be said that the mentioned combination of DR techniques is able to recover part of the information related to the spatial geometry which was not captured by the feature geometry, and do so in a low-dimensional space.

Table 9: CC scores for t-SNE with Isomap with number of neighbors $K_{\text{Iso}} = \{8, 12\}$ as initial solution, using the mentioned dissimilarity features

d_{diss}	K_{Iso}	KS	K							
			10		25		50		100	
			TW	CT	TW	CT	TW	CT	TW	CT
CMD	8	0.360	0.941	0.948	0.938	0.947	0.935	0.946	0.930	0.941
	12	0.361	0.941	0.948	0.939	0.947	0.935	0.946	0.930	0.942
$\mathbb{C}\{\cdot\}$	8	0.324	0.941	0.937	0.939	0.936	0.936	0.935	0.932	0.932
	12	0.325	0.942	0.938	0.938	0.938	0.935	0.936	0.931	0.934
S. sub.	8	0.405	0.945	0.951	0.942	0.949	0.939	0.948	0.934	0.943
	12	0.409	0.945	0.949	0.942	0.948	0.939	0.947	0.934	0.943

4.1.1 Use of non-labeled data for CC

Related to what has been discussed in section 3.4.8, about the use of unlabeled data for CC, the results of this approach have not been satisfactory. The quality scores of the charts described in this section—using the labeled data only—is, in general, greater than those in which a part of the unlabeled data has been included, up to 6000 unlabeled data points. Therefore, the addition of unlabeled data for chart construction has not been further explored in this thesis, focusing only on charts learned from labeled locations.

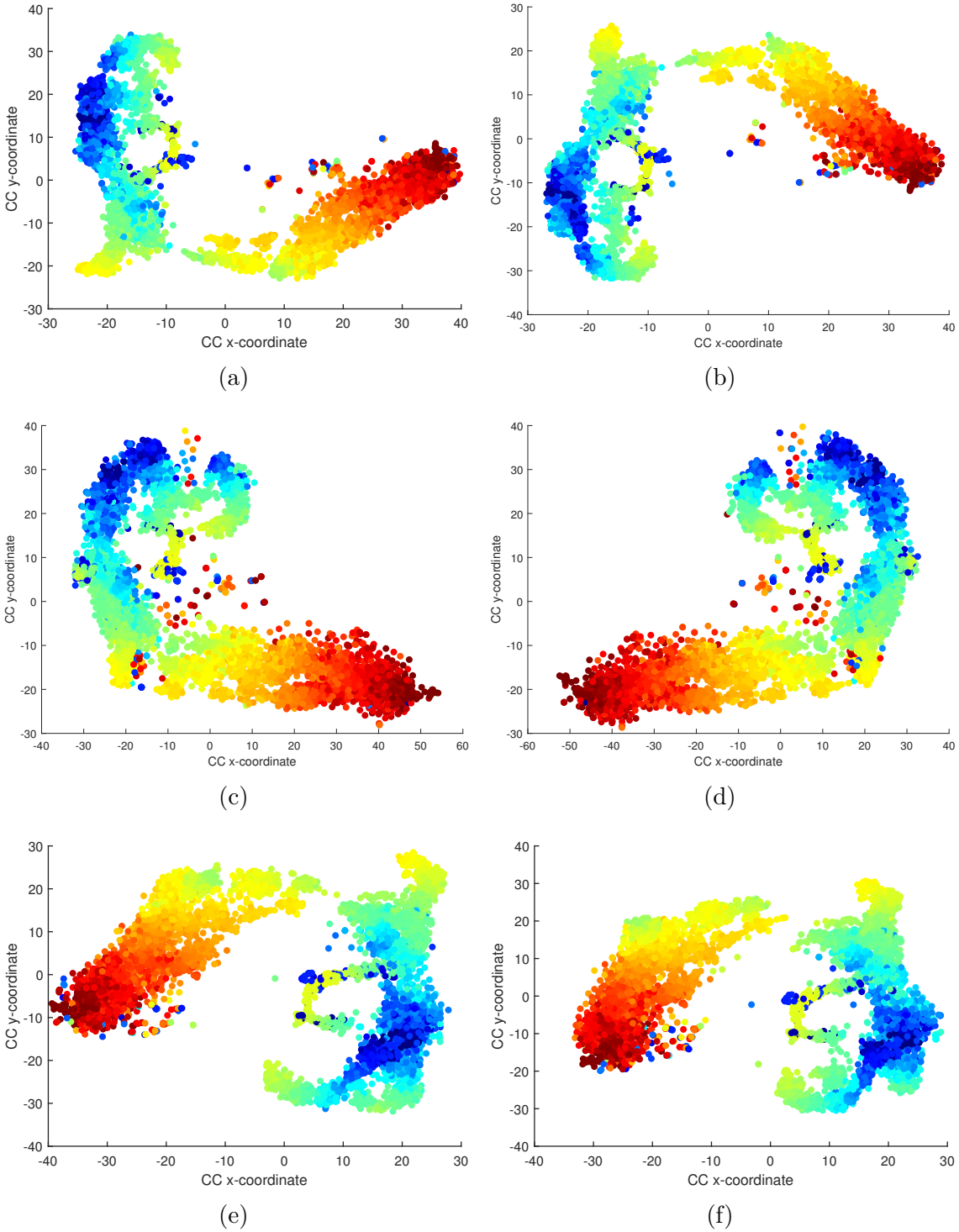


Figure 13: Charts obtained from applying t-SNE with Isomap (with $p' = 2$) as initial solution, for the mentioned dissimilarities and parameters: (a); CMD, Isomap: $K = 8$. (b); CMD, Isomap: $K = 12$. (c); C$\{\cdot\}$, Isomap: $K = 8$. (d); C$\{\cdot\}$, Isomap: $K = 12$. (e); S. sub., Isomap: $K = 8$. (f); S. sub., Isomap: $K = 12$. The value of *Perp* for t-SNE is 200 for all charts, the rest of the parameters are the same ones used on the charts described on Table 8

4.2 Localization results

In this section, the results related to localization are presented. Firstly, the focus lies on localization by using the dissimilarity features directly. These results are obtained from the localization techniques presented in section 3.6. After that, results concerning the use of the charts in figure 13 are presented, followed by a comparison of the mentioned results.

4.2.1 Results related to KNN methods

This section presents the results related to performing localization by utilizing the different KNN methods introduced in section 3.6.1. These results are presented as follows: for each dissimilarity based feature, the mean localization error in [m] (39) is shown, rounded to the nearest integer. The size of the test set is 25% of the total of points. The number of neighbors, K, is chosen to be the optimal in each case.

On Table 10, the localization results corresponding to dissimilarity based features, before using them for CC, are presented.

Table 10: Mean localization error in [m], obtained using the KNN methods. DV: Distance vector. IDV: Inverse Distance Vector

d_{diss}	KNN method		
	Pairwise	DV	IDV ($\alpha = 1$)
CMD	44	43	43
Chordal	52	54	54
X	51	55	54
$\mathbb{C}\{\cdot\}$	45	44	43
S. Sub.	42	42	40

4.2.2 Results related to the Extreme Learning Machine

The results concerning localization with the use of an ELM, as introduced in section 3.6.2 can be found on Table 11. From the results, it can be observed that the performance of ELM is poorer than that of the KNN methods previously detailed. This situation is analogous to that occurring in [33], and as a consequence, the result shows that ELMs are, in general, a worse localization algorithm than all KNN methods presented. The fact that the ELM shows a poorer performance might be due to having a shallow structure, with only one hidden layer, which might not be sufficient to form a function which is able to correctly map features to locations. However, the convenience of ELMs lies in the fact that they are relatively simple structures in which a relatively small number of weights is learned. Moving to larger neural network frameworks might give more accurate localization results, but with the drawback of an increased complexity which would reduce the efficiency and defeat the purpose of simplicity motivating ELMs. Moreover, deeper networks require larger datasets, which is not the case with the used dataset.

Table 11: Mean localization error in [m], obtained using the ELM for the optimal values of number of neurons μ and regularization term γ

d_{diss}	Avg. loc. error in [m]	μ	γ
CMD	58	700	10^{-1}
Chordal	76	1300	10^{-2}
X	65	700	10^{-2}
$\mathbb{C}\{\cdot\}$	59	700	10^{-2}
S. Sub.	60	800	10^{-2}

4.2.3 Results of localization based on CC

As discussed in section 3.6.3, channel charts can be used for the purpose of localization. It is worthwhile to observe the performance of the localization methods on the charts shown on Figure 13, as they are the charts that have the highest scores. Table 12 shows the average localization error when using the distances in the charts, for the three KNN methods and for the ELM. The test set consists of 25% of the total number of points. These localization results are worse than those of dissimilarity features, portrayed on Tables 10 and 11.

Table 12: Mean localization error in [m], obtained from the charts in figure 13. DV: Distance Vector. IDV: Inverse Distance Vector. The regularization term for the ELM $\gamma = 10^{-1}$

d_{diss}	K_{Iso}	KNN			ELM	
		Pairwise	DV	IDV ($\alpha = 1$)	Avg. error	μ
CMD	8	49	49	49	54	600
	12	48	48	49	52	800
$\mathbb{C}\{\cdot\}$	8	49	48	49	54	800
	12	49	49	50	54	800
S.Sub.	8	46	46	47	53	800
	12	46	46	50	52	500

4.2.4 Comparison between KNN and ELM

It is important to compare not only the average localization error for the KNN methods and ELM, as the previous sections have discussed, but also to look at the distribution of errors. This can be done by observing the empirical CDF of the localization methods in question. This result can be found in Figure 14, for the CMD, for a test set of 25% of the total number of transmitters. The picture includes the localization error distributions for localization using dissimilarity and channel charts. From the picture it is clear that the three KNN methods yield the same test error distribution, and the performance of ELM is worse than that of the KNN methods. For KNN, it can be seen that around 50% of estimations have an error

less than 25 m, and the larger average of 43-44 m is due to locations with significant errors, caused by false neighbors in the feature geometry. The relatively high value of trustworthiness for the CMD, found on Table 7, explains this result; a significant amount of transmitters can be located with a relatively small localization error, and they have high point-wise TW scores (see section 3.2.2). For some points, there exist neighbors in the feature geometry which are not present in the original spatial geometry—and their point-wise TW is low—, and for those points the location estimation will be far from the true location.

It is important to discuss the results concerning the three KNN methods. It has been found that there are practically no differences among them, when it comes to the localization errors. This can be said for both the localization based on dissimilarity features and based on charts. That means, that the dissimilarity or chart distance between two transmitters is as accurate as the distance between the dissimilarity vectors. Another consequence of this, and despite it has not been presented in the results, is the fact that the use of weighting, as detailed in equation (38), does not change the localization results in this case. From the perspective of computation, and since the pairwise KNN approach is the quickest to compute, it is sensible to use it, to the detriment of the distance vector approaches, since in this case they practically yield the same results. Additionally, and since weighting has little effect on the results, it is sensible not to use it, since it is the most sensible option computationally-wise.

It can be concluded that localization based on CC, for the dataset in question, has a poorer performance than if dissimilarity vectors are used directly. Despite the charts obtained with t-SNE from Isomap as initial solutions being a better representation—and having the advantage of having significantly less dimensions—than the dissimilarity features, i.e., the quality scores are higher than those of the features, in this case this fact does not translate to localization performance. This might be due, in part, to the fact that t-SNE represents probability, not distances. Therefore, the obtained CC cannot be used to infer the distances between transmitters, and localization results suffer from this fact.

4.3 Graph signal processing

The main application of the graph interpolation algorithm presented in section 3.5 is to estimate the unknown locations of transmitters, having information on the true locations of a set of those transmitters. With this, one can calculate the localization error of the estimations produced by the algorithm with equation (39), analogously to the other localization methods.

In this thesis, the localization results of this method are evaluated with a training set of 3000 points, i.e., $|\mathcal{S}| = 3000$. The following parameters are chosen as follows: $T = 0.1, \theta^2 = 0.5$ (see Algorithm 7). Figure 15 illustrates the estimations of the points in the test set, when using data points containing dissimilarities. Table 13 includes the average localization errors of the unlabeled locations, and the quality scores related to each of the charts, which are now constructed in a semi-supervised manner. Note that the referred quality scores are computed only taking into account the unlabeled points, as including the labeled points into the computation would

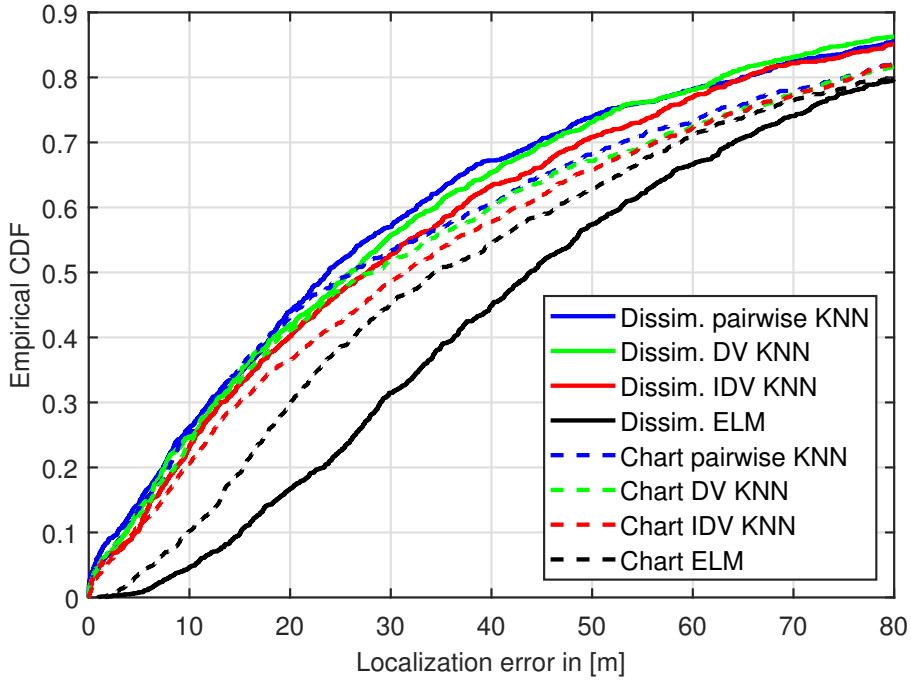


Figure 14: Empirical CDF of the localization error of a test set consisting of 25% of all labeled data points, corresponding to the KNN methods and the ELM, for the CMD. It can be seen that the ELM performs worse than the KNN methods, which practically have the same error distribution. The solid curves represent the error when performing localization on the dissimilarity features (eq. (42)), while the dashed curves are for localization based on CC, with the particular chart being from t-SNE with Isomap with $K_{\text{Iso}} = 8$ as initial solution performed on the CMD dissimilarity features, as shown in Figure 13a.

essentially divide the scores by a constant related to the relative sizes of the training and test sets, since the locations of the labeled points are known, and only those of the transmitters in the unabeled set are estimated.

It can be seen that the KS corresponding the unlabeled points is lower than that of the other charts previously discussed. This can be explained from understanding the essence of the graph interpolation method. Here, the predicted locations, for the unlabeled points, are estimated, thus attempting to preserve the distances. Unsupervised CC does not have true location information, and as a consequence the distance distortion—measured by the KS—will, in general, be worse than that of charts constructed with semi-supervised approaches.

Another aspect worth mentioning is the fact that, despite the charts being of relatively high quality, in this case the location estimations are worse than those found with the dissimilarity features (for the KNN methods) and based on CC, on Tables 10 and 12, respectively.

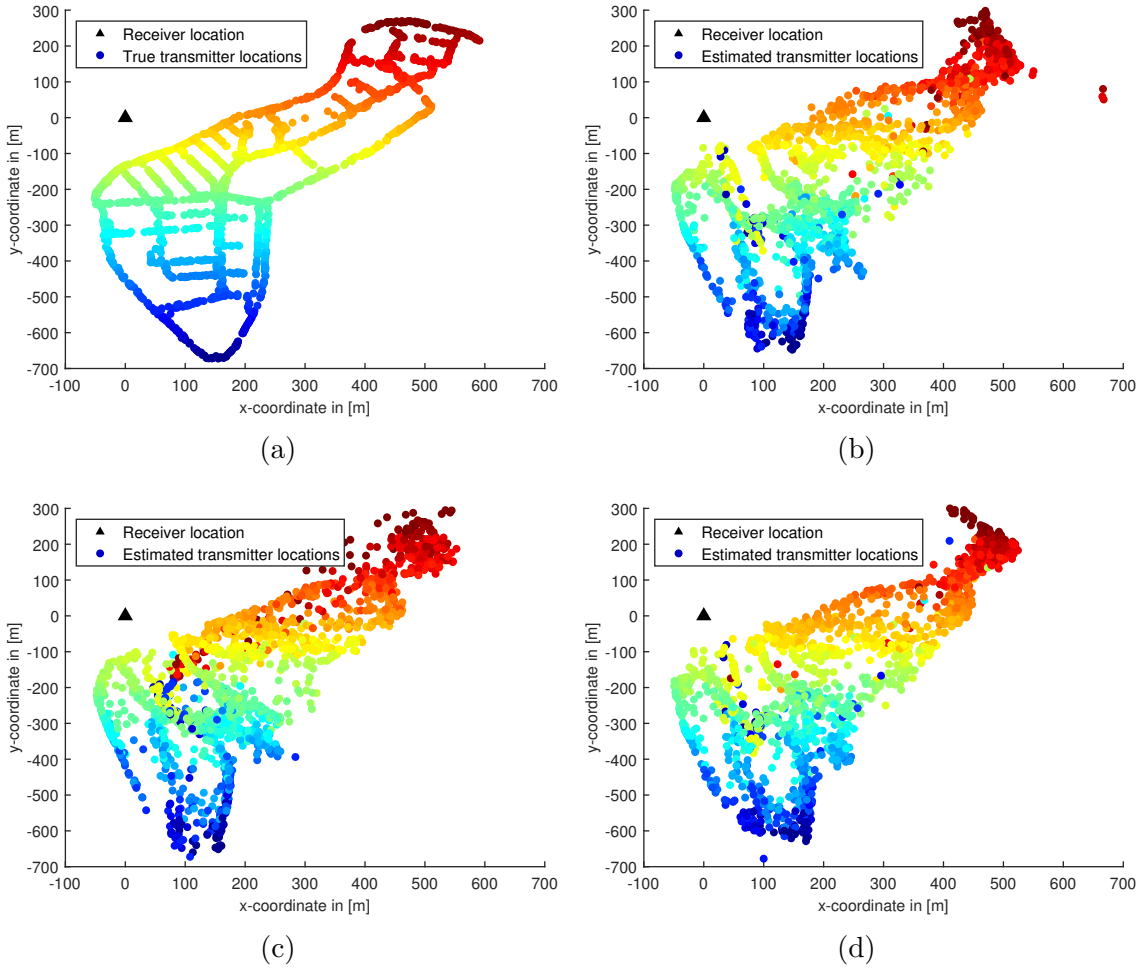


Figure 15: Location estimations of the points in the test set after the graph interpolation algorithm. (a) True locations of the unlabeled points. (b) Estimations when $d_{\text{diss}} = \text{CMD}$. (c) Estimations when $d_{\text{diss}} = \mathbb{C}\{\cdot\}$. (d) Estimations when d_{diss} is the signal subspace related dissimilarity.

Table 13: Quality scores and localization errors of the estimations produced by the graph interpolation algorithm, for the given dissimilarities

d_{diss}	KS	$K = 50$		Avg. loc. error in [m]
		TW	CT	
CMD	0.2097	0.9387	0.9460	51
$\mathbb{C}\{\cdot\}$	0.2740	0.9227	0.9253	61
S. sub.	0.2197	0.9343	0.9417	54

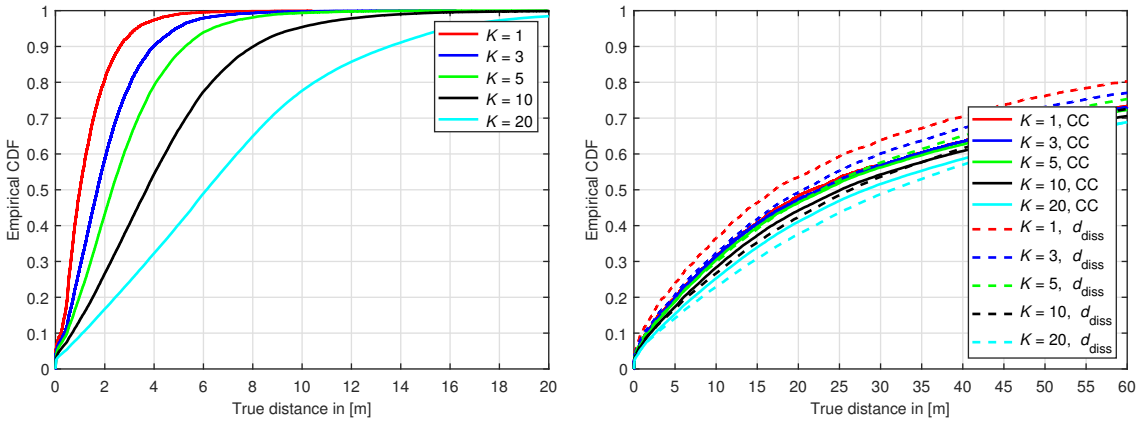
4.4 Effect of the spatial sampling on the localization results

It is sensible to look at the localization results from the perspective of the spatial density of the transmitters, i.e., the true pairwise distances between transmitters, which conform the spatial geometry. This way, the localization results can be easier

understood. Note that the obtained localization accuracies have no intrinsic meaning, since they depend on the density of the transmitters, i.e., in their number, N, the size of the radio cell, and how uniformly the users are spread about the cell.

One aspect worth analyzing is the statistical properties of the distances between transmitters and their K-nearest neighbors. The comparison can be made by defining the distances used to find the neighbors. On one hand, the true neighbors can be analyzed. Secondly, the neighbors can be chosen from the CC distances, as it is done in CC based localization. Finally, they can also be chosen based on the dissimilarity features already discussed. Figure 16 illustrates these concepts, by analyzing the empirical CDF of the true distances between transmitters and their neighbors, for different values of K-nearest neighbors. It can be seen that there is a significant difference between the distances related to the true neighbors and those related to the neighbors in the channel chart and those related to a dissimilarity feature. The KNN based localization is based on this phenomenon, and depends on the distances of the neighbors to make the location estimations. It can be observed that the closest true neighbors are located at relatively low distances, which confirms that there exist neighboring locations in the channel chart which are not neighbors in the spatial geometry, which rises the average localization error. It is important to notice that, in Figure 16, the plots involve all transmitters, whereas only the true locations in the training set (in this thesis, 75% of the total number of transmitters) are used for the estimations. However, this fact has a smaller effect than the presence of false neighbors in the channel charts, due to the fact that if the average location error is in the order of 45-50 m, for some transmitters in the test set their estimated location is far from their true location.

Table 14 shows the average and the median of the distances between transmitters and their K-nearest neighbors, for different values of K-nearest neighbors. The comparison is made between the true neighbors, the neighbors found on the charts and the neighbors related to dissimilarities. There is a clear difference between the true neighbors and the other two definitions, as has already been seen in Figure 16. There is a small difference between the true distances to the nearest neighbors for the channel chart and dissimilarity distances, and they are higher than those related to the true neighbors. It is important to remark the fact that there is a high difference between the average and the median of the true distances when the neighbors are found from a channel chart and a dissimilarity. The values of the TW detailed on Tables 7 and 9 mean that there exist false neighbors—relatively few, as the average values are higher than 0.9—in the feature geometry and in the channel chart. However, the TW value does not give any information about the true distances of transmitters to those false neighbors. By observing these relatively large differences between the mean and the median of the distances to the nearest neighbors, it can be further concluded that a non-negligible number of transmitters have neighbors in the feature geometry as well as in the channel charts which raise the mean localization error at a considerable amount.



(a) (b)

Figure 16: Empirical CDF of the true distances between transmitters and their K-nearest neighbors. (a) True neighbors. (b) Neighbors found from the chart obtained from t-SNE with Isomap ($K_{\text{Iso}} = 8$) as initial solution, for the CMD (Fig. 13a), and those found in the feature geometry, for the CMD features.

Table 14: Mean and median of the true distances in [m] between transmitters and their K-nearest neighbors, considering the true neighbors, the neighbors found from the chart obtained from t-SNE with Isomap ($K_{\text{Iso}} = 8$) as initial solution, for the CMD (Fig. 13a), and the neighbors found directly with the CMD features

K	True neighbors		CC neighbors		d_{diss} neighbors	
	Mean	Median	Mean	Median	Mean	Median
1	1.27	0.99	54.66	21.66	43.78	17.06
3	2.00	1.70	55.38	22.66	49.45	20.61
5	2.66	2.29	57.56	23.51	52.12	22.82
10	4.15	3.69	59.79	25.15	56.66	26.35
20	6.84	6.13	62.50	28.15	61.53	31.38

5 Future work

There is a significant number of future work directions that have not been covered in this thesis. It is patent that the localization results are in need to be improved, as the physical separation between transmitters is much lower than the average localization errors found. A possible direction would consist of processing the CSI, using suitable signal processing techniques, in order to find out the distribution of the antenna array of the receiver. This information can be then used for extracting more complex features which capture the nature of the antenna array. In addition, use other more complex feature extraction techniques that might more accurately capture the geometry in which the UEs reside.

Since feature extraction and DR are, in general, problem specific, it is not possible to know how the specific dissimilarities and techniques would perform on other datasets. A logical continuation of this thesis would be to use different datasets and be able to draw more generalized conclusions on the performance of the methods and spot potential improvements and challenges. Following this line of thought, a deeper mathematical analysis of feature extraction and DR algorithms might be needed, in order to understand the aspects of CSI are relevant for the construction of accurate channel charts and features. Also, it could provide an insight on what features and DR techniques are suitable for CSI of different properties.

There exist other DR algorithms that have not been utilized in this thesis. Thus, a natural future work direction would be to analyze the results of charts constructed with other DR techniques. Additionally, and since unlabeled data is available, construction of charts using part of this data can lead to improvements on the quality of charts and localization results.

The question of utilizing other semi-supervised techniques for learning channel charts remains open. The graph interpolation technique used in this thesis constitutes a way of generating charts in a semi-supervised manner, but there might be alternative methods for creating such charts. Moreover, it is important to note, that the algorithm utilized in this thesis (shown in Algorithm 7) has the same limitations as the other DR techniques, in terms of the quality of the high-dimensional features that they take as input. A natural research avenue would be to investigate the effect of feature properties on the semi-supervised generated charts, in order to provide the algorithm with new features which give better results.

Metric learning approaches with various degrees of supervision are an interesting potential future work concept. This way, the dissimilarities can not only be calculated directly from the extracted features, but a learning framework is used in addition, which can allow for features based on learned metrics or dissimilarities.

6 Conclusions

This thesis has described and analyzed methods to construct channel charts and perform user localization, from real—not simulated—CSI collected at a multi-antenna BS from transmitting UEs scattered about an outdoor area approximately covering 0.5 km^2.

There is an important number of conclusions and statements that arise from the results obtained in this thesis, whose mention is relevant for the understanding of the results, and gives enlightenment to potential further analysis emerging from this work.

One of the most relevant ideas that can be derived is the fact that creating channel charts with the desired properties for this particular dataset has proven to be a challenging task. Charts which are unable to accurately represent the true locations of the transmitters are of little use. Therefore, many of the charts that have been generated throughout this thesis work, such as those whose scores are shown on Table 8 are unsatisfactorily suited for localization applications. This phenomenon has several reasons which, in part, boil down to a few particularities about the CSI data studied in this thesis. First of all, and as has been previously discussed, the dataset in question has not been in any way created specifically for the prospect of generating charts. Secondly, the uncertainty related to the locations of the malfunctioning antennas introduces an additional challenge. Aside from these facts, it is patent that the CSI data have properties which are undesired, from the point of view of extracting features. One of the most impactful of these properties is the existence of two highly distinct sets of transmitters, namely the LoS and NLoS sets. As has been extensively discussed in chapter 3, the existence of transmitters for which most of the recorded CSI values are 0 limits the effective information that can be gained from the CSI and, consequently, makes features more prone to suffer from considerable inaccuracies. In addition, the possible non-negligible amount of noise present in the data hinders the possibility of having features of a higher quality.

As a consequence of the described circumstances, utilizing the generated charts for the prospect of localization can lead to challenges in improving the results of localization performed directly on the features, thus making channel charts unsuitable to be used for this purpose, even though their quality scores are an improvement to those of the features extracted from the CSI, as has been the case in this thesis.

The semi-supervised technique of graph signal processing, detailed in section 3.5 has proven to be able to provide relatively good results in terms of the quality of the charts obtained. However, localization based on these charts was poorer than the other methods, namely KNN and ELM based on dissimilarity features and on charts, as high and low-dimension representations of the spatial geometry, respectively.

In conclusion, the results presented in chapter 4 are remarkable in that, it has been possible to generate radio maps with good quality, despite the numerous challenges present, and despite some choices that have been taken to solve the original problem. This work has the focus on the analysis of CC and localization with real data. Despite the numerous limitations that the dataset utilized in this thesis has, as have been thoroughly explained, it is remarkable that it has been possible to obtain good quality

charts, when compared to charts learned from synthetic data in [3, 10, 11, 22], where the antenna structure is known and there is room for adapting some parameters to the specific needs of the problem. The generated charts can give good results for other applications, including Radio Resource Management (RRM), but are in need of improvement if they are to be used for localization.

As mentioned in chapter 5, there are many open questions and research directions to explore, which can lead to interesting findings and results.

References

[1] K. Chandra, R. V. Prasad, B. Quang, and I. Niemegeers, "CogCell: cognitive interplay between 60 GHz picocells and 2.4/5 GHz hotspots in the 5G era," *IEEE Communications Magazine*, vol. 53, no. 7, pp. 118–125, 2015.

[2] J. Deng, O. Tirkkonen, R. Freij-Hollanti, T. Chen, and N. Nikaein, "Resource allocation and interference management for opportunistic relaying in integrated mmWave/sub-6 GHz 5G networks," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 94–101, 2017.

[3] C. Studer, S. Medjkouh, E. Gönültaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[4] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 41–53, 2005.

[5] N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich, and M. Coulon, "Direct localization for massive MIMO," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2475–2487, 2017.

[6] R. Talmon, I. Cohen, and S. Gannot, "Supervised source localization using diffusion kernels," in Proc. of *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011, pp. 245–248.

[7] Y. Chapre, A. Ignjatovic, A. Seneviratne, and S. Jha, "CSI-MIMO: Indoor Wi-Fi fingerprinting system," in Proc. of *39th Annual IEEE Conference on Local Computer Networks*, 2014, pp. 202–209.

[8] V. Savic and E. G. Larsson, "Fingerprinting-based positioning in distributed massive MIMO systems," in Proc. of *IEEE 82nd Vehicular Technology Conference (VTC)*, 2015, pp. 1–5.

[9] J. Vieira, E. Leitinger, M. Sarajlic, X. Li, and F. Tufvesson, "Deep convolutional neural networks for massive MIMO fingerprint-based positioning," in Proc. of *IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–6.

[10] P. Ferrand, A. Decurninge, and M. Guillaud, "DNN-based localization from channel estimates: Feature design and experimental results," *arXiv preprint arXiv:2004.00363*, 2020.

[11] P. Ferrand, A. Decurninge, L. G. Ordoñez, and M. Guillaud, "Triplet-based wireless channel charting," *arXiv preprint arXiv:2005.12242*, 2020.

[12] A. Sobehy, E. Renault, and P. Muhlethaler, "NDR: Noise and Dimensionality Reduction of CSI for indoor positioning using deep learning," in Proc. of *IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[13] I. S. Dhillon, J. R. Heath, T. Strohmer, and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *Experimental Mathematics*, vol. 17, no. 1, pp. 9–35, 2008.

[14] R. Hajizadeh, A. Aghagolzadeh, and M. Ezoji, "Local distances preserving based manifold learning," *Expert Systems with Applications*, vol. 139, 2020.

[15] A. Schwartz and R. Talmon, "Intrinsic isometric manifold learning with application to localization," *SIAM Journal on Imaging Sciences*, vol. 12, no. 3, pp. 1347–1391, 2019.

[16] E. Lei, O. Castañeda, O. Tirkkonen, T. Goldstein, and C. Studer, "Siamese neural networks for wireless positioning and channel charting," in Proc. of *57th Annual Allerton Conference on Communication, Control, and Computing*, 2019, pp. 200–207.

[17] P. Agostini, Z. Utkovski, and S. Stańczak, "Channel charting: an euclidean distance matrix completion perspective," in Proc. of *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5010–5014.

[18] P. Huang, O. Castañeda, E. Gönültaş, S. Medjkouh, O. Tirkkonen, T. Goldstein, and C. Studer, "Improving channel charting with representation-constrained autoencoders," in Proc. of *IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2019, pp. 1–5.

[19] E. Kupershtein and M. Wax, "Indoor localization based on multipath fingerprinting," Master's thesis, Technion-Israel Institute of Technology, Faculty of Electrical Engineering, 2013.

[20] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[21] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors. I. the MUSIC algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1758–1774, 1992.

[22] J. Deng, S. Medjkouh, N. Malm, O. Tirkkonen, and C. Studer, "Multipoint channel charting for wireless networks," in Proc. of *52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 286–290.

[23] "IEEE CTW 2020 data competition," 2020, accessed: 8-10-2020. [Online]. Available: http://ctw2020.ieee-ctw.org/wp-content/uploads/sites/94/2020/02/CTW2020-Localization-Competition.pdf

[24] H. Al-Tous, P. Kazemi, and O. Tirkkonen, "Coordinated uplink precoding for spatially consistent mmWave channel covariance measurements," in Proc. of *IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.

[25] M. Arnold, J. Hoydis, and S. ten Brink, "Novel massive MIMO channel sounding data applied to deep learning-based indoor positioning," in Proc. of *12th International ITG Conference on Systems, Communications and Coding*. VDE, 2019, pp. 1–6.

[26] X. Dang, J. Ren, Z. Hao, Y. Hei, X. Tang, and Y. Yan, "A novel indoor localization method using passive phase difference fingerprinting based on channel state information," *International Journal of Distributed Sensor Networks*, vol. 15, no. 4, 2019.

[27] J. Venna and S. Kaski, "Neighborhood preservation in nonlinear projection methods: An experimental study," in Proc. of *International Conference on Artificial Neural Networks*, 2001, pp. 485–491.

[28] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[29] H.-U. Bauer and K. R. Pawelzik, "Quantifying the neighborhood preservation of self-organizing feature maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 4, pp. 570–579, 1992.

[30] J. A. Lee and M. Verleysen, "Quality assessment of dimensionality reduction: Rank-based criteria," *Neurocomputing*, vol. 72, no. 7-9, pp. 1431–1443, 2009.

[31] M. Herdin, N. Czink, H. Ozcelik, and E. Bonek, "Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels," in Proc. of *IEEE 61st Vehicular Technology Conference (VTC)*, 2005, pp. 136–140.

[32] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 876–890, 2014.

[33] A. Decurninge, L. G. Ordóñez, P. Ferrand, H. Gaoning, L. Bojie, Z. Wei, and M. Guillaud, "CSI-based outdoor localization for massive MIMO: Experiments with a learning approach," in Proc. of *15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–6.

[34] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: a comparative review," Tilburg University, Technical Report TiCC-TR 2009-005, 2009.

[35] L. Van der Maaten, "An introduction to dimensionality reduction using Matlab," Maastricht University, Technical Report MICC 07-07, 2007.

[36] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[37] F. Wickelmaier, "An introduction to MDS," *Sound Quality Research Unit, Aalborg University, Denmark*, vol. 46, no. 5, pp. 1–26, 2003.

[38] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in Proc. of *Advances in Neural Information Processing Systems*, 2002, pp. 585–591.

[39] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[40] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[41] L. K. Saul and S. T. Roweis, "Think globally, fit locally: unsupervised learning of low dimensional manifolds," *Journal of Machine Learning Research*, vol. 4, no. Jun, pp. 119–155, 2003.

[42] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.

[43] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[44] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in Proc. of *Advances in Neural Information Processing Systems*, 2003, pp. 857–864.

[45] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in Proc. of *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 5445–5449.

[46] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.